How Much Should Portfolios Shrink?*

Chulwoo Han[†]

This Version: June 2019

Abstract

This paper develops a portfolio model that penalizes the deviation from a reference portfolio. The proposed model renders a robust portfolio that performs superior under parameter uncertainty. Penalizing the deviation also improves the performance of existing shrinkage portfolio models that are sub-optimal due to model parameter uncertainty. The equal-weight portfolio turns out to be a better reference portfolio than the currently holding portfolio even in the presence of transaction costs. A data-driven method for determining the degree of penalization is offered. Comprehensive simulation and empirical studies suggest that the proposed model significantly outperforms various existing models.

JEL Classification: G11

Keywords: Optimal portfolio; deviation penalty; shrinkage estimator; parameter uncertainty; estimation error

*I would like to thank Bing Han (Editor), and an anonymous referee, whose comments and suggestions have significantly improved the paper. I thank Raymond Kan, Victor DeMiguel, and seminar participants at Hanyang University, KAIST, and University of Stavanger for their helpful discussions and comments. Any errors are my own.

[†]I can be reached at: Department of Finance and Economics, Durham Business School, Mill Hill Lane, Durham, DH1 3LB, UK; chulwoo.han@durham.ac.uk.

1 Introduction

Optimal portfolio choice under parameter uncertainty and transaction costs is a central problem in the portfolio literature, and there has been a significant amount of effort dedicated to this problem.

One pillar has been formed by the Bayesian approach: *e.g.*, Klein and Bawa (1976); Brown (1976, 1978); Jorion (1986); Black and Litterman (1992); Pástor (2000); Pástor and Stambaugh (2000), among others. For a review of Bayesian models, the reader is referred to Avramov and Zhou (2010). More recently, the robust optimization that optimizes portfolio under a worst-case scenario becomes popular: *e.g.*, Goldfarb and Iyengar (2003); Fabozzi et al. (2007); Cao et al. (2009); Ceria and Stubbs (2016). Kan and Zhou (2007) and Tu and Zhou (2011) optimally combine two or more portfolios so that the expected utility loss is minimized. Incorporating transaction costs is also known to help reduce the sensitivity and improve the performance after transaction costs: *e.g.*, Gârleanu and Pedersen (2013); DeMiguel et al. (2015). Other approaches impose weight constraints (Jagannathan and Ma, 2003) or use a shrinkage method for parameter estimation (Ledoit and Wolf, 2004).

While these models are known to alleviate the problems arising from parameter uncertainty and perform superior to the classical mean-variance model, DeMiguel et al. (2009) show that none of the portfolio models considered in their paper consistently outperforms the naïve, equal-weight portfolio. Their work triggered many studies that challenge the equal-weight portfolio: *e.g.*, Tu and Zhou (2011); Kirby and Ostdiek (2012); Bessler et al. (2014). Their evaluation method that compares risky-asset-only portfolios derived from optimal portfolios has also been criticized for being unfair to some models (see, *e.g.*, Kirby and Ostdiek (2012) and Kan et al. (2016)). Still, most optimal strategies seem to struggle to outperform the naïve strategy consistently across assets and time.

This paper addresses parameter uncertainty by developing a portfolio model that penalizes the deviation from a reference portfolio (deviation penalty, henceforth), where the reference portfolio can be any portfolio known at the time of rebalancing such as the current portfolio or the equal-weight portfolio. By penalizing the deviation from a reference portfolio, the model reduces the time-series variation of portfolio weights and generates a portfolio that is robust to estimation errors. While the model can be considered a shrinkage model, existing shrinkage models can also benefit from the deviation penalty as illustrated below.

This paper offers a data-driven method to determine the degree of penalization and shows that the proposed model significantly outperforms many existing models in terms of certainty equivalent and Sharpe ratio before and after transaction costs. The optimal degree of penalization turns out to be strikingly high compared to the shrinkage levels of other shrinkage models, especially when the input parameters are subject to large estimation errors.

Comprehensive simulation and empirical studies involving thirteen datasets and a sample period of over 60 years suggest that the proposed model performs superior in comparison to various existing models such as the equal-weight portfolio, market portfolio, and other shrinkage models. Robustness tests show that the proposed model continues to outperform other models in different circumstances: *e.g.*, during different sample periods, during recession periods, and when using different sample sizes for input parameter estimation.

Apart from introducing a new portfolio model that performs superior, the paper makes several important contributions to the extant literature.

One important contribution of the paper is to show that, both theoretically and empirically, the equal-weight portfolio is a more effective reference portfolio than the current portfolio. While penalizing the deviation from the current portfolio is helpful to some extent especially when trades are subject to transaction costs, its effect on portfolio performance is revealed to be rather trivial compared to the effect of penalizing the deviation from the equal-weight portfolio: shrinking towards the equal-weight portfolio renders a less volatile portfolio with better performance. Counterintuitively, the equal-weight portfolio also incurs fewer transaction costs. Using the current portfolio is certainly a more effective way of reducing turnover for a single period. However, since the current portfolio can be distant from the true optimal portfolio under parameter uncertainty, shrinking towards it is revealed to cause higher turnover and transaction costs in the long run. As penalizing the deviation from the current portfolio is similar to accounting for transaction costs, this finding is contrary to the earlier findings that accounting for transaction costs in portfolio optimization enhances portfolio robustness and performance: *e.g.*, Gârleanu and Pedersen (2013); DeMiguel et al. (2015); Olivares-Nadal and DeMiguel (2018).

Another important contribution of the paper is to show that existing shrinkage models, *e.g.*, Kan and Zhou (2007) and Tu and Zhou (2011), are sub-optimal and can be improved substantially when augmented with the deviation penalty. These models combine two or more portfolios so that the expected out-of-sample utility is maximized. However, the coefficients on the portfolios (model parameters) are nonlinear functions of unknown input parameters and therefore inherit their uncertainty, resulting in worse-than-expected performance even when the underlying assumptions are correct. That is, just like the mean-variance optimal portfolio is not optimal when the mean and covariance matrix are subject to estimation errors, the shrinkage models are not optimal when the model parameters are subject to estimation errors inherited from the mean and covariance matrix. If any of the assumptions such as *i.i.d.* normal returns are violated, which is very likely, the problem is exacerbated. Although the impact of model parameter uncertainty can be substantial, the extant literature has failed to recognise this. This paper shows that the deviation penalty increases the robustness of existing shrinkage portfolios and alleviates the performance deterioration due to model parameter uncertainty.

The rest of the paper is organized as follows. Section 2 develops portfolio models with the deviation penalty. A calibration method to determine the degree of penalization is also offered here. Section 3 describes the datasets and portfolio models used in the empirical study. Section 4 evaluates the proposed models via simulations. Two reference portfolios, the equal-weight and the current portfolios, are examined. Section 5 carries out empirical studies which compare the proposed models against various existing models. Section 6 concludes the paper. The implementation details of the models used in the empirical analysis and the full empirical results are provided in the accompanying internet appendix (Internet Appendix).

2 Optimal Portfolio with Deviation Penalty

2.1 Utility Maximization

The quadratic utility maximization problem with the deviation penalty is given by

$$\max_w U(w) = w'\mu - \frac{\gamma}{2}w'\Sigma w - \frac{\delta}{2}(w - w_0)'G(w - w_0), \tag{1}$$

where $\mu \in \mathbb{R}^N$ and $\Sigma \in \mathbb{R}^{N\times N}$ are the mean and covariance matrix of N asset returns in excess of the risk-free rate, $w \in \mathbb{R}^N$ is the portfolio weights, and γ is the risk aversion coefficient of the investor.[1] The last term on the right-hand side penalizes the deviation from a reference portfolio w_0, where δ is a constant and $G \in \mathbb{R}^{N\times N}$ is a penalty matrix. The reference portfolio w_0 can be any portfolio known at the time of portfolio rebalancing: the equal-weight portfolio, w_{ew}, and the current portfolio, w_{t-}, are considered in this paper.

The optimal portfolio w^* that maximizes the utility function is given by

$$w^* = (\gamma\Sigma + \delta G)^{-1}(\mu + \delta G w_0). \tag{2}$$

If an asset return has a large variance, its mean estimate may well have a large estimation error, and it is justifiable to penalize the weight change of such assets more severely. From this perspective, a natural choice of G would be the covariance matrix, Σ. When $G \equiv \Sigma$, the optimal portfolio becomes a convex combination of the Markowitz (1952) optimal portfolio, $w_{ml} = \frac{1}{\gamma}\Sigma^{-1}\mu$, and the reference portfolio, w_0:

$$w^* = \frac{\gamma}{\gamma + \delta}w_{ml} + \frac{\delta}{\gamma + \delta}w_0. \tag{3}$$

[1] Returns refer to excess returns throughout the paper unless otherwise noted.

In order to implement w^*, unknown μ and Σ need to be estimated. If the asset returns are *i.i.d.* normal random variables, the maximum likelihood (ML) estimates of μ and Σ, $\hat{\mu}$ and $\hat{\Sigma}$, are independent of each other and have the following distributions:

$$\hat{\mu} \sim \mathcal{N}\left(\mu, \frac{\Sigma}{T}\right), \quad \hat{\Sigma} \sim \mathcal{W}_N(T-1, \Sigma)\frac{1}{T}, \tag{4}$$

where T is the estimation window size, and $\mathcal{N}$ and $\mathcal{W}_N$ respectively denote a normal distribution and N-dimensional Wishart distribution. To allow the case when asset returns are not *i.i.d.* or $\hat{\mu}$ and $\hat{\Sigma}$ are estimated separately, *e.g.*, using different estimation windows, a slightly relaxed assumption,

$$\hat{\mu} \sim \mathcal{N}\left(\mu, \frac{\Sigma}{K}\right), \quad \hat{\Sigma} \sim \mathcal{W}_N(T-1, \Sigma)\frac{1}{T}, \tag{5}$$

for some constant K, is made.

An unbiased estimate of the Markowitz portfolio is then given by

$$\hat{w}_{ml} = \frac{1}{\gamma}\tilde{\Sigma}^{-1}\hat{\mu}, \quad \tilde{\Sigma} = \frac{T}{T-N-2}\hat{\Sigma}. \tag{6}$$

As shown by Kan and Zhou (2007), however, plugging $\hat{w}_{ml}$ in (3) is not optimal in the sense that it does not minimize the expected utility loss. Therefore, the following generic form

$$w(a,b) = a\hat{w}_{ml} + bw_0 \tag{7}$$

is considered, and a and b are determined so that the expected utility loss is minimized, or equivalently, the expected out-of-sample utility (expected utility, henceforth) is maximized:

$$\max_{a,b} E[U(a,b)] = E\left[w(a,b)\mu - \frac{\gamma}{2}w(a,b)'\Sigma w(a,b) - \frac{\delta}{2}(w(a,b)-w_0)'\Sigma(w(a,b)-w_0)\right]. \tag{8}$$

Proposition 1. *The optimal a and b that solve* (8) *are given by*

$$a^* = \frac{\gamma}{\gamma+\delta} a_0^*, \tag{9}$$

$$b^* = \frac{\gamma}{\gamma+\delta} b_0^* + \frac{\delta}{\gamma+\delta}, \tag{10}$$

where

$$a_0^* = \frac{\theta^2 - \psi^2}{c_1\left(\frac{N}{K}+\theta^2\right) - \psi^2}, \tag{11}$$

$$b_0^* = \frac{c_1\left(\frac{N}{K}+\theta^2\right) - \theta^2}{c_1\left(\frac{N}{K}+\theta^2\right) - \psi^2} \frac{1}{\gamma} \frac{w_0'\mu}{w_0'\Sigma w_0}, \tag{12}$$

$$\theta^2 = \mu'\Sigma^{-1}\mu, \quad \psi^2 = \mu_0'\Sigma^{-1}\mu, \quad \mu_0 = \frac{w_0'\mu}{w_0'\Sigma w_0}\Sigma w_0, \tag{13}$$

and

$$c_1 = \frac{(T-2)(T-N-2)}{(T-N-1)(T-N-4)}. \tag{14}$$

See Appendix A.1 for proof. Note that variables a_0^* and b_0^* are the optimal a and b when $\delta = 0$. Since a^* and b^* are functions of unknown μ and Σ, they need to be estimated, and one method is provided in Appendix A.2.

The optimal portfolio is given by

$$\begin{aligned} w(a^*, b^*) &= a^*\hat{w}_{ml} + b^* w_0 \\ &= \frac{\gamma}{\gamma+\delta}\left(a_0^*\hat{w}_{ml} + b_0^* w_0\right) + \frac{\delta}{\gamma+\delta} w_0 \\ &= \frac{\gamma}{\gamma+\delta}\left(a_0^*\hat{w}_{ml} + (1-a_0^*) w_{im}\right) + \frac{\delta}{\gamma+\delta} w_0, \end{aligned} \tag{15}$$

where

$$w_{im} = \frac{1}{\gamma}\frac{w_0'\mu}{w_0'\Sigma w_0} w_0 = \frac{1}{\gamma}\Sigma^{-1}\mu_0. \tag{16}$$

The portfolio w_{im} is proportional to w_0 and can be interpreted as the Markowitz portfolio when the mean returns are μ_0.

When $w_0 = w_{ew}$ and $\delta = 0$, the optimal portfolio becomes the shrinkage portfolio of Tu and Zhou (2011) except that they set $b = 1-a$. There is no reason to assume $b = 1-a$ apart from the obvious advantage of having fewer parameters. Furthermore, under this restriction, the proportion of $\hat{w}_{ml}$ to w_0 is no longer invariant to γ.

Viewed as a function of δ, $w^*(\delta) = w(a^*, b^*|\delta)$, the optimal portfolio can be rewritten as

$$w^*(\delta) = \frac{\gamma}{\gamma+\delta} w^* + \frac{\delta}{\gamma+\delta} w_0, \tag{17}$$

where $w^* = a_0^* \hat{w}_{ml} + b_0^* w_0$ is the solution to the usual expected utility maximization problem without the deviation penalty term. In fact, any shrinkage estimator of the form, $w(a, b) = a\hat{w} + bw_0$ for some portfolio $\hat{w}$, has the optimal solution given in (17) with $w^* = a_0^* \hat{w} + b_0^* w_0$ being the optimal solution when $\delta = 0$ (a_0^* and b_0^* here are generic notations to denote the optimal values and not as defined in (11) and (12)).

2.2 Variance Minimization

The penalty term can also be incorporated into a variance minimization problem:

$$\begin{aligned} \min_w \; & V(w) = \frac{1}{2} w' \Sigma w + \frac{\delta}{2} (w - w_0)' \Sigma (w - w_0) \\ \text{subject to } & w' 1_N = 1, \end{aligned} \tag{18}$$

where $1_N \in \mathbb{R}^N$ is a vector of ones, and $w_0' 1_N = 1$ is assumed. The optimal portfolio that solves (18) is given by

$$w^* = \frac{1}{1+\delta} w_{mv} + \frac{\delta}{1+\delta} w_0, \tag{19}$$

where $w_{mv} = \frac{\Sigma^{-1} 1_N}{1_N' \Sigma^{-1} 1_N}$ is the global minimum-variance portfolio. An unbiased estimate of w_{mv} can be obtained from

$$\hat{w}_{mv} = \frac{\hat{\Sigma}^{-1} 1_N}{1_N' \hat{\Sigma}^{-1} 1_N}. \tag{20}$$

As before, a generic portfolio strategy

$$w(a) = a\hat{w}_{mv} + (1-a)w_0 \tag{21}$$

is considered and a is found so that the expected variance is minimized:

$$\min_a E[V(a)] = E\left[\frac{1}{2}w(a)'\Sigma w(a) + \frac{\delta}{2}(w(a)-w_0)'\Sigma(w(a)-w_0)\right]. \tag{22}$$

Since $\hat{w}_{mv}'1_N = 1$ and $w_0'1_N = 1$, the budget constraint is implicitly satisfied without further restriction.

Proposition 2. *The optimal a that solves (22) is given by*

$$a^* = \frac{1}{1+\delta}\frac{\sigma_0^2 - \sigma_{mv}^2}{\sigma_0^2 - \left(1 - \frac{N-3}{T-N+1}\right)\sigma_{mv}^2}, \tag{23}$$

where $\sigma_0^2 = w_0'\Sigma w_0$ and $\sigma_{mv}^2 = w_{mv}'\Sigma w_{mv} = (1_N'\Sigma^{-1}1_N)^{-1}$ are the variances of w_0 and w_{mv}, respectively.

See Appendix B for proof and the estimation of a^*.

2.3 Choice of Reference Portfolio

The choice of the reference portfolio has a significant impact on portfolio performance. As opposed to the conventional wisdom, the current portfolio does not appear to be an effective shrinkage target and is usually dominated by the equal-weight portfolio. The latter renders far less volatile portfolios and involves lower transaction costs, leading to robust performance especially under high parameter uncertainty and transaction costs. This is shown theoretically by the following propositions and confirmed by the simulations and empirical studies in the following sections.

Proposition 3. *Suppose that portfolio w is rebalanced t times via a deviation penalty model.*

Let w_t^c and w_t^e denote the optimal portfolio at time t respectively when $w_0 = w_{t-}$ and $w_0 = w_{ew}$:

$$\begin{aligned} w_t^c &= (1-\alpha)w_t^* + \alpha w_{t-1}^c, \\ w_t^e &= (1-\alpha)w_t^* + \alpha w_{ew}, \end{aligned} \tag{24}$$

where w_t^* is the optimal portfolio of the base model, and $\alpha = \dfrac{\delta}{\gamma+\delta}$. At time 0, the expectation and variance of the optimal portfolio at time t are given by:

$$\begin{aligned} E(w_t^c) &= (1-\alpha^t)E(w_t^*) + \alpha^t w, \\ V(w_{it}^c) &= (1-\alpha)^2 V(w_{it}^* + \alpha w_{it-1}^* + \cdots + \alpha^{t-1} w_{i1}^*), \\ E(w_t^e) &= (1-\alpha)E(w_t^*) + \alpha w_{ew}, \\ V(w_{it}^e) &= (1-\alpha)^2 V(w_{it}^*), \end{aligned} \tag{25}$$

where w_{it}^c and w_{it}^e are the i-th elements of w_t^c and w_t^e, respectively. It follows that

$$E(w_t^c) \to E(w_t^*) \text{ as } t \to \infty, \tag{26}$$

$$V(w_{it}^e) < V(w_{it}^c) < V(w_{it}^*). \tag{27}$$

See Appendix C.1 for proof. The proposition implies that while w_t^e is biased, *i.e.*, $E(w_t^e) \neq E(w_t^*)$, $V(w_{it}^e)$ can be considerably smaller than $V(w_{it}^*)$ resulting in superior performance especially under high parameter uncertainty. Empirical studies show that α is often greater than 0.5, in which case, $V(w_{it}^e) < 0.25V(w_{it}^*)$. On the contrary, although w_t^c is asymptotically unbiased, $V(w_{it}^c)$ can be considerably larger than $V(w_{it}^e)$ especially when t is large, resulting in an inferior performance of w_t^c.

The effect of penalization is more pronounced when transaction costs are taken into account. As opposed to our intuition, penalizing the deviation from w_{ew} incurs lower transaction costs than w_{t-} in the long run, as illustrated in Proposition 4.

Proposition 4. *Let* $\Delta w_{it} = w_{it} - w_{it-1}$.[2] *If* $E\left[\Delta w_{it}^{*} \Delta w_{it-1}^{c}\right] > -\frac{\alpha}{2(1-\alpha)} E\left[(\Delta w_{it-1}^{c})^2\right]$, *the following inequalities hold:*

$$E\left[(\Delta w_{it}^{e})^2\right] < E\left[(\Delta w_{it}^{c})^2\right] < E\left[(\Delta w_{it}^{*})^2\right]. \tag{28}$$

See Appendix C.2 for proof. $E\left[\Delta w_{it}^{*} \Delta w_{it-1}^{c}\right] > -\frac{\alpha}{2(1-\alpha)} E\left[(\Delta w_{it-1}^{c})^2\right]$ is a reasonable assumption as the correlation between Δw_{it}^{*} and Δw_{it-1}^{c} is usually small and negligible when the estimation window is large. If they are uncorrelated, the following relationship can be established.

Corollary 1. *If* $E\left[\Delta w_{it}^{*} \Delta w_{it-1}^{c}\right] = 0$,

$$E\left[(\Delta w_{it}^{e})^2\right] < (1-\alpha^2) E\left[(\Delta w_{it}^{c})^2\right]. \tag{29}$$

Proposition 4 implies that both reference portfolios reduce transaction costs, but w_{ew} is more effective especially when α is large.

2.4 Calibration of δ

The coefficient δ needs to be determined in order to implement the model. This section proposes a simple, but effective calibration method. The procedure is as follows:

1. For the first ten months into the sample period, δ is set to 3, 2, or 1 respectively when $T = 60$, 120, or 240.

2. When $t > 10$, δ is calibrated each month so that the certainty equivalent (CE) during $1, \ldots, t-1$ is maximized. The optimal δ is found via line search spanning the range $[0, 10]$.

[2]For simplicity, turnover is defined as $w_{it} - w_{it-1}$ instead of $w_{it} - w_{it-} = w_{it} - (1 + r_{it-1}) w_{it-1}$. In the latter case, the inequalities hold under a slightly more complex assumption.

3. In the presence of transaction costs, δ is chosen based on the CE after transaction costs.

2.5 Optimal Portfolio Choice under Constraints

One drawback of the proposed model is that it is difficult to extend it to a constrained optimization problem, *e.g.*, utility maximization with short-sale constraints. This is also true for other shrinkages models. This paper employs a simple method to incorporate constraints as described below. Given an optimal portfolio, $\hat{w}^*$, obtained from a deviation penalty model without constraints, the expected returns implied by the portfolio can be derived as follows:

$$\bar{\mu} = \gamma \hat{\Sigma} \hat{w}^*. \tag{30}$$

A constrained problem is then solved as usual after substituting $\hat{\mu}$ with $\bar{\mu}$.

Even though this method does not explicitly maximize the expected utility subject to constraints, empirical studies suggest that this method effectively accounts for parameter uncertainty in a constrained problem. The same approach can be adopted for a constrained variance minimization problem.

3 Data and Portfolio Models

3.1 The Data

The deviation penalty models are evaluated on the thirteen datasets described in Table I and compared against the portfolio models listed in Table III. The datasets are based on those used in DeMiguel et al. (2009), Kirby and Ostdiek (2012), and Kan et al. (2016), but also include new ones. Except for the first dataset D1 which has the sample period from 1990.10 to 2015.12, all other datasets have the same sample period from 1951.01 to 2015.12. The sample period refers to the out-of-sample period during which portfolios are rebalanced and evaluated, and the samples for moments estimation extend further to the past. For example, when $T = 240$, the mean and covariance matrix of the asset returns in the first month are estimated using the sample from 1931.01 to 1950.12. By using the same out-of-sample period regardless of the estimation window size, the results from different estimation window sizes ($T = 60$, 120, and 240 months in this paper) can be directly compared. The moments of the asset returns are estimated monthly during the evaluation period rolling the estimation window.

[TABLE I]

Before assessing the performance of the portfolio models, it is worth understanding the characteristics of the datasets. Table II reports a summary of the *ex-post* optimal portfolios, *i.e.*, Markowitz portfolios obtained from the mean and covariance matrix of the entire sample. As evidenced by the sum of the absolute values of the weights (the fifth column), the *ex-post* optimal portfolios are unrealistically highly leveraged in most datasets. They even short the risky portfolio in D6 and D7. As shown in the last three columns, these datasets also frequently yield negative expected returns on the global minimum-variance portfolio, which will lead to a short risky portfolio position in the optimal portfolio.[3] High leverage arises largely from the inclusion of the market and factor portfolios: these portfolios can be approximated by other assets and the optimal portfolio often resembles a long-short strategy (sell the market and buy other assets). Without the factor portfolios, the datasets behave more nicely resulting in less leveraged portfolios. Nevertheless, the empirical studies of this paper are primarily based on the datasets including factor portfolios as these have been used in previous studies, *e.g.*, DeMiguel et al. (2009). The results from the datasets without factor portfolios are provided in the Internet Appendix.

[TABLE II]

[3] There is a marked contrast between D5 which contains only the market portfolio and D6 and D7 which contain all three Fama-French factors: adding the Fama-French factors results in higher leverage and short positions of the risky portfolio.

3.2 The Portfolio Models

Table III lists the portfolio models that are compared in this paper. The *ex-post* optimal portfolio (W*) is the Markowitz portfolio obtained from the sample moments of the entire sample. The equal-weight portfolio (EW) is chosen as a benchmark, and other classical portfolio strategies, *i.e.*, the Markowitz optimal portfolio (ML), global minimum-variance portfolio (MV), and their short-sale constrained versions (ML+, MV+) are also considered. The models (VT, OC) of Kirby and Ostdiek (2012) are added as they are argued to outperform EW. The three-fund rule (KZ) of Kan and Zhou (2007) and the shrinkage portfolios (TZML, TZKZ) of Tu and Zhou (2011) are included as they share the same approach to parameter uncertainty and are similar to the proposed models without the deviation penalty term.

The deviation penalty models (DPML, DPMV) are tested using two reference portfolios, w_{ew} and w_{t-}. In addition, a model (DPMLK) that estimates K in (5) instead of assuming $K = T$ is examined. K is estimated using the method of Lo and MacKinlay (1988) as described in Appendix A.3.

Variants of KZ, TZML, and TZKZ that incorporate the deviation penalty are also considered. If $w_0 = w_{ew}$, the optimal Tu and Zhou portfolio incorporating the deviation penalty is given by

$$w_{tz}(\delta) = \frac{\gamma}{\gamma+\delta} w_{tz} + \frac{\delta}{\gamma+\delta} w_0, \tag{31}$$

where w_{tz} is the original Tu and Zhou portfolio (TZML or TZKZ). Extension of KZ is less straightforward. In principle, it would be best to determine the coefficients on the three portfolios simultaneously by letting

$$w(a, b, c) = a\hat{w}_{ml} + b\hat{w}_{mv} + cw_0, \tag{32}$$

and determining a, b, and c so that the expected utility is maximized. The solution to this problem is given in the Internet Appendix. Although this approach is theoretically more appealing, it is difficult to implement due to the complexity of model parameter estimation: simulations reveal that a crude plug-in method using the ML estimates, $\hat{\mu}$ and $\hat{\Sigma}$, performs unsatisfactorily. Instead, based on the generic solution in (17), the following form is employed:

$$w_{kz}(\delta) = \frac{\gamma}{\gamma + \delta} w_{kz} + \frac{\delta}{\gamma + \delta} w_0, \tag{33}$$

where w_{kz} is the original Kan and Zhou three-fund rule. Equation (31) and (33) are also used when $w_0 = w_{t-}$. Implementation details of each model can be found in the Internet Appendix.

[TABLE III]

4 Simulation Studies

The proposed models are first validated via simulation studies. Four datasets, D1, D2, D5, and D8, out of the thirteen datasets in Table I are chosen for simulation. The sample mean and covariance matrix of the entire sample are regarded as the true mean and covariance matrix. The expected utility and variance are obtained from 10,000 iterations. These values are computed without the penalty term as portfolio performance should not be affected by it.

4.1 Utility Maximization

Table IV reports normalized expected utilities for the case of $\gamma = 3$. The first column represents the portfolio models, and the numbers in the header are estimation window sizes. EW and EW* are equal-weight portfolios adjusted so as to maximize utility, respectively using the sample and true moments. The reported values are averages across the datasets. Detailed results from each dataset can be found in the Internet Appendix.

[TABLE IV]

Consistent with the findings in the literature, *e.g.*, Tu and Zhou (2011) and Kan et al. (2016), ML is outperformed by EW when the window size is small: $T > 240$ is required for ML to outperform EW. KZ improves over ML significantly and outperforms both EW and ML across all window sizes. It is however outperformed by the two models of Tu and Zhou (2011). Of the two, TZKZ performs superior.

Compared with TZML, DPML(0) performs marginally, but consistently superior. It also has smaller standard errors. This result is in favour of the proposed two-parameter model over the one-parameter model of Tu and Zhou (2011). In fact, the difference becomes more prominent when $\gamma = 1$ (available in the Internet Appendix).

Of particular interest is the effect of penalization, which can be examined by comparing the DPML models with different values of δ. Performance enhancement stemming from penalization is substantial when T is small: the expected utility of DPML(0) is 0.200 when $T = 60$, whereas those of DPML(1) and DPML(2) are 0.286 and 0.312, respectively. These values are even higher than the expected utility of EW* which assumes prior knowledge of the distribution. This result is rather striking considering that DPML(δ), $\delta > 0$, is a linear combination of DPML(0) and the equal-weight portfolio. Furthermore, the fact that incorporating the penalty causes the objective utility function to drift away from the evaluation utility function makes the result particularly noteworthy. Due to this misalignment, the δ associated with the maximum utility declines as T increases, *i.e.*, as parameter uncertainty diminishes, and the models with $\delta > 0$ eventually underperform DPML(0). This result suggests that a carefully chosen δ for a given level of parameter uncertainty would improve portfolio performance. This hypothesis is further investigated later in this section.

The penalty term does not only enhance the expected utility but also reduces its standard error considerably: *e.g.*, when $T = 60$, the standard error of DPML(0) is 0.507, whilst those of DPML(1) and DPML(2) are 0.274 and 0.185, respectively. A smaller standard error implies a smaller chance of extreme losses over a finite investment horizon.

Jagannathan and Ma (2003) show that imposing short-sale constraints can reduce estimation error even when the constraints are wrong. A similar conclusion can be drawn here. ML+, DPML+(0), and DPML+(1) all exhibit superior performance to ML when $T < 360$. This result is remarkable considering the high leverage of the *ex-post* optimal portfolios reported in Table II. The performance of DPML+ is particularly noteworthy. It outperforms ML+ significantly and outperforms EW* across all T's. Besides, it has significantly lower standard errors compared to ML+. This confirms that the proposed method of incorporating constraints is effective. Nevertheless, a limitation of short-sale constrained models is that their performance is suppressed even when T is large.

Unlike simulation, the real-world performance of optimal portfolios does not necessarily improve with the estimation window size. DeMiguel et al. (2009) find that accumulating the estimation window rather than rolling it improves the performance of optimal models only slightly, whereas no apparent relationship between performance and window size can be derived from the empirical results in Tu and Zhou (2011). It appears that beyond a certain window size, parameter uncertainty does not diminish further or even rises again. From this perspective, the proposed models that show robust performance when T is small are expected to demonstrate superior performance when applied to actual market data.

4.2 Sensitivity to Misspecification

The assumption that the returns are *i.i.d.* random variables is rather strong and unrealistic, and the expectation of the sample mean may well deviate from the true mean. In order to examine the effect of misspecification, the mean is perturbed when random samples are drawn using $\mu_i := \mu_i(1 + 0.2z_i)$, where μ_i is the mean return of asset i and z_i is a standard normal random variable. Simulation results are reported in Table V.

[TABLE V]

It is striking how small errors in mean can deteriorate the performance of shrinkage estimators: KZ, TZML, TZKZ, and DPML(0) all perform poorly and yield negative utilities regardless of the size of T. In fact, the performance of these models worsens with T. This is because these models ignore the misspecification and put more weight on ML as T increases. This may explain to some extent why some shrinkage estimators perform worse when the estimation window size is larger: see, *e.g.*, Table 6 of Tu and Zhou (2011). On the contrary, the deviation penalty model is far more robust to misspecification. DPML($\delta > 0$) maintains positive expected utility and its performance improves with T. Furthermore, it has considerably smaller standard errors. The short-sale-constrained models are also robust to misspecification of mean.

Accounting for parameter uncertainty does help improve portfolio performance. KZ, TZML, and TZKZ all improve over ML and generally outperform EW even for a moderately large T. However, since their model parameters need to be estimated, these models are still sensitive to estimation errors and misspecification, and their actual performance can be unexpectedly poor. Meanwhile, the deviation penalty models are robust to misspecification and demonstrate superior performance when subject to large estimation errors. In addition, they have much smaller standard errors.

4.2.1 Performance over a Finite Investment Horizon

The results in Table IV and V are asymptotic properties. A real-world investment horizon is finite and the performance of portfolios can be different. To examine the performance over a finite investment horizon, portfolios are assumed to be managed for ten years during which they are rebalanced monthly. As we now have the "current portfolio", the DPML models with $w_0 = w_{t-}$ are also examined.

[TABLE VI]

Table VI reports the mean and standard deviation of the normalized certainty equivalents (CE) obtained from 10,000 iterations. The results in the upper (lower) panel are before (after) transaction costs. Transaction costs of 30 basis points (bp) for both buying and selling risky assets and 0 bp for the risk-free asset are assumed.

The mean certainty equivalents are similar to the expected utilities in Table IV and will not be discussed further. What is more interesting is a comparison between the DPML models with $w_0 = w_{ew}$ (DPML(δ)) and those with $w_0 = w_{t-}$ (DPMLc(δ)). When $T > 60$, DPMLc(δ) has a higher CE than its counterpart. Adding the current portfolio has an effect similar to accumulating estimation sample and DPMLc(δ) is anticipated to outperform DPML(δ) under the *i.i.d.* assumption. Notwithstanding, it underperforms DPML(δ) when $T = 60$ both before and after transaction costs and has a higher standard error. This is because when the moments are subject to large estimation errors, the current portfolio can be distant from the true optimal portfolio and it becomes a less effective shrinkage target compared to a fixed-weight portfolio. As illustrated in the next section, the equal-weight portfolio indeed serves better as the reference portfolio.

4.3 Variance Minimization

Table VII reports the results from variance minimization. The top panel reports expected variances and the bottom panel reports sample variances from the finite investment horizon.

The standard global minimum-variance portfolio (MV) performs well even for a small T: its expected variance is 27.6% higher than the *ex-post* optimal value when $T = 60$ and only 11.5% higher when $T = 120$. Still, DPMV(0) yields consistently lower variances across all window sizes. DPMV(0) also has smaller standard errors. Meanwhile, incorporating deviation penalty with $\delta = 1$ adds little value. This is perhaps because the estimation error of the covariance matrix is not large enough to benefit from deviation penalty. Both DPMVc(0) and DPMVc(1) perform comparably to DPMV(0).

[TABLE VII]

4.4 Model Parameter Uncertainty and Optimal Deviation Penalty

The results so far suggest that the deviation penalty does improve the performance of optimal portfolios, and the optimal degree of penalization decreases as estimation window size increases and therefore parameter uncertainty diminishes. This section investigates the optimal degree of penalization more in detail and demonstrates that other shrinkage models can also be improved by incorporating the deviation penalty.

As shown in Section 2, the coefficients of the portfolios in a shrinkage model (model parameters) are a function of unknown input parameters and need to be estimated. This model parameter uncertainty can deteriorate portfolio performance significantly as reported in Table VIII.

[TABLE VIII]

The table compares the certainty equivalents obtained from true model parameters ('True') with those from estimated model parameters ('Estd'). The utility loss due to model parameter uncertainty is substantial especially when the input parameters are subject to large estimation errors. For instance, when $T = 60$, The certainty equivalent of KZ is reduced from 0.176 to 0.058 and that of TZML is reduced from 0.360 to 0.184. Although the loss diminishes rapidly with T, it remains significant even when $T = 240$. This implies that the coefficient on the shrinkage target determined by maximizing expected utility is not sufficient as model parameter uncertainty is ignored. As illustrated in the rest of this section, the deviation penalty helps compensate this loss.

Figure 1 displays the certainty equivalent of each model as a function of δ in the absence of transaction costs. Solid lines are for $w_0 = w_{ew}$ and dashed lines are for $w_0 = w_{t-}$. Adding the penalty term improves the performance of the models in most cases, and its effect is particularly noticeable when $w_0 = w_{ew}$ and T is small: *e.g.*, the CE of KZ increases by 20% at $\delta \approx 1.6$ when $T = 120$, while it continues to increase with δ within the considered range when $T = 60$. Among the four models, only KZ does not contain w_{ew} in its original form and thus benefits most by incorporating w_{ew}, whereas TZKZ, which includes both w_{ew} and w_{kz}, benefits least. In contrast to the case of $w_0 = w_{ew}$, penalizing the deviation from w_{t-} has little effect on performance when there is no transaction cost.[4] This result was predicted in Proposition 3 where it is shown that the expected portfolio weights of the models that penalize the deviation from w_{t-} converge to those of their base models, whilst their variances are not particularly smaller.

[FIGURE 1]

The effect of the deviation penalty is more pronounced when transaction costs are taken into account. The results in Figure 2 are obtained assuming transaction costs of 30 bp. In the presence of transaction costs, the models with $w_0 = w_{ew}$ as well as those with $w_0 = w_{t-}$ improve portfolio performance, but the improvement is more prominent when $w_0 = w_{ew}$. The optimal δ that maximizes CE is strikingly large often exceeding the considered range: recall that, with $\gamma = 3$, $\delta = 3$ implies 50% on the reference portfolio. When $w_0 = w_{ew}$, the actual loading on w_{ew} is even higher than what δ implies as TZML, TZKZ, and DPML already involve w_{ew} without the penalty term. As illustrated in Proposition 4, anchoring to w_{ew} incurs lower transaction costs in the long run and is therefore more effective than anchoring to w_{t-} even in the presence of transaction costs: the gap between the two versions indeed widens with transaction costs.

[FIGURE 2]

DPMLc is different from other models with $w_0 = w_{t-}$ in that it consists of only $\hat{w}_{ml}$ and w_{t-} and their loadings are dynamically determined based on the input parameters, whereas the other models consist of the base model and w_{t-} and their loadings are solely determined by δ. Dynamically determining the loadings results in the superior performance of DPMLc to DPML when the *i.i.d.* assumption holds and estimation errors are small. However, as

[4] An exception to this is DPML, which is discussed further later in this section.

revealed in the case of $T = 60$, DPMLc can be outperformed by DPML when subject to high uncertainty. In addition, DPML's performance improves faster with δ. Since $\hat{w}_{ml}$, w_{t-}, and their loadings are all subject to estimation errors, DPMLc is very sensitive and its performance can rapidly deteriorate under large estimation errors. As will be seen later in the empirical studies, DPMLc indeed performs very poorly when applied to real market data.

As opposed to the conventional wisdom, the current portfolio does not appear to be an effective shrinkage target and is usually dominated by the equal-weight portfolio. The latter renders far less volatile portfolios and involves lower transaction costs, leading to robust performance especially under high uncertainty and transaction costs.

The deviation penalty models can be extended by incorporating both w_{ew} and w_{t-}:

$$w^*(\delta, \kappa) = (1 - \kappa) \left(\frac{\gamma}{\gamma + \delta} w^* + \frac{\delta}{\gamma + \delta} w_0 \right) + \kappa w_{t-}. \tag{34}$$

Figure 3 presents simulation results from this extension. Solid lines represent the case of $\kappa = 0$ and are the same as those in Figure 2. Incorporating both w_{ew} and w_{t-} further improves performance especially when T is small. Nevertheless, the gain from the inclusion of w_{t-} is not large and diminishes with δ.

[FIGURE 3]

5 Empirical Studies

5.1 Portfolio Construction and Evaluation

Portfolios are rebalanced every month during the sample period based on the mean and covariance estimates obtained from a rolling estimation window of size $T = 60$, 120, or 240. Monthly portfolio returns and performance measures are then calculated.

It is nontrivial to compare different portfolio models on a level playing field. DeMiguel et al. (2009) compare the risky portfolios derived from the optimal portfolios by normalizing the risky asset weights. This method, however, gives an unfair disadvantage to some models that are not designed to maximize the Sharpe ratio. This issue is discussed in detail in Kan et al. (2016) for the Kan and Zhou model. Another, more subtle and often overlooked problem is that if the optimal weight of the risky portfolio (the sum of the risky asset weights) is negative, the maximum Sharpe ratio portfolio does not exist and the naïve scaling whereby the weights are divided by their sum leads to the minimum Sharpe ratio portfolio. In this case, the original portfolio and the derived risky portfolio have opposite exposures to the risky assets and their performances will differ considerably.[5] As shown in Table II, negative risky portfolio weights are indeed common especially when T is small or in D6 and D7.

Comparing utility maximizing portfolios has its own problem as the results depend on the choice of the risk aversion coefficient. Besides, for those portfolio strategies that do not take the mean into account, *e.g.*, EW, MV, and VT, adjusting the weights so that the utility is maximized conflicts their nature as the adjustment involves the mean.

In this paper, portfolios are constrained so that they have the same *ex-ante* variance (variance targeting). This creates a more level playing field for evaluation without favouring a particular model. Imposing a risk constraint is also common in practical asset allocation. If the *ex-ante* variance of an optimal portfolio is $\hat{\sigma}_p^2$ and the target variance is σ_{max}^2, the constraint can be satisfied by scaling the portfolio weights with $\sigma_{max}/\hat{\sigma}_p$.[6] In the empirical studies, the target variance is defined as the variance of the equal-weight portfolio over the entire sample period. Kirby and Ostdiek (2012) adjust portfolios so that they have the same expected return as the equal-weight portfolio. Whilst this method is similar to variance targeting, the results will be less reliable due to the sizeable estimation error in mean. Imposing constraints on the mean is also less common in practice.

As a robustness test, utility maximizing portfolios are also compared. In this case, the

[5]For this reason, DeMiguel et al. (2009) normalize portfolio weights by the absolute value of their sum. In this case, however, the normalized portfolio still includes the risk-free asset.

[6]If the optimal portfolio has a short position on the risky portfolio, it is scaled by $-\sigma_{max}/\hat{\sigma}_p$.

portfolios that are not designed to maximize utility, *i.e.*, EW, MV, VT, TMV, and DPMV, are compared without adjusting the weights so as to maximize utility.[7] These models are revealed to perform better unadjusted.

Portfolios are evaluated using four performance measures: certainty equivalent (CE), Sharpe ratio (SR), turnover (TO), and leverage (LV). These are defined as follows:

$$\mathrm{CE} = \bar{r}_p - \frac{\gamma}{2} s_p^2; \tag{35}$$

$$\mathrm{SR} = \frac{\bar{r}_p}{s_p}; \tag{36}$$

$$\mathrm{TO} = \frac{1}{NT_o} \sum_{t=1}^{T_o} \sum_{i=1}^{N} |w_{i,t} - w_{i,t-}|; \tag{37}$$

$$\mathrm{LV} = \frac{1}{T_o} \sum_{t=1}^{T_o} \sum_{i=1}^{N} |w_{i,t}|, \tag{38}$$

where $\bar{r}_p$ and s_p are the mean and standard deviation of the portfolio returns over the sample period, T_o is the sample size, and $w_{i,t-}$ and $w_{i,t}$ are the weights of asset i immediately before and after rebalancing at time t. For certainty equivalent, $\gamma = 3$ is used.

To assess the effect of transaction costs, SR and CE are calculated both before and after transaction costs assuming transaction costs of 30 bp for buying and selling risky assets and 0 bp for the risk-free asset. LV is calculated to gauge the feasibility of portfolios. Highly leveraged portfolios are not desirable and unrealistic to many investors. While imposing constraints on asset weights can yield more realistic portfolios, its impact on portfolio performance will be nontrivial, making the performance attributable to the unique characteristics of individual models indistinguishable. Therefore, portfolios are evaluated without weight constraints, and LV is calculated to supplement the results.

Given thirteen datasets and four performance measures, a coherent evaluation and ranking of the models is a formidable task. To facilitate evaluation, SR and CE are normalized by the SR and CE of the *ex-post* optimal portfolio. Normalization helps measure the loss

[7] One exception is DPMV+. DPMV+ solves the usual utility maximization problem employing the mean implied by the unconstrained DPMV.

against the hypothetical maximum and allows us to compare performance across datasets. The normalized measures are averaged across the datasets in order to generate a single measure for each performance metric. Despite the fact that the summary statistics across datasets depend on the choice of datasets, this provides a convenient way of comparing models. The empirical analysis in this section is primarily based on the summary statistics, and the results from each dataset are referred to when necessary.

The empirical results are reported in Table IX (variance targeting) and Table X (utility maximization). In each table, the column 'Mean CE' reports the mean of the normalized CE across the datasets, and 'N(>EW)', referred to as outperformance ratio, reports the number of datasets in which a model outperforms EW in terms of CE. The Internet Appendix provides full results including; results on other performance measures, results from datasets without factor portfolios, and dataset-level results.

[TABLE IX]

[TABLE X]

5.2 Main Findings

The most remarkable finding from the empirical studies is the performance enhancement resulting from the deviation penalty. Consistent with the simulation results, incorporating the deviation penalty improves portfolio performance significantly, not only for the proposed models, DPML and DPMV, but also for the existing ones; KZ, TZML, and TZKZ. Recall that these models already address parameter uncertainty before incorporating the deviation penalty. In general, a model involving three portfolios (KZ and TZKZ) outperforms a model involving two portfolios (TZML and DPML), and KZ is revealed to perform best overall when augmented with the deviation penalty.

Another important finding is the sharp contrast between the shrinkage targets, w_{ew} and w_{t-}. As anticipated from the simulation studies, incorporating w_{t-} is not as effective as incorporating w_{ew}, both before and after transaction costs. When shrunk towards w_{t-}, KZ, TZML, and TZKZ do not perform better than their base models before transaction costs. Their performances improve after transaction costs, but the effect is limited especially under utility maximization. The results from DPMLc is particularly noteworthy. Contrary to the simulation results where DPMLc performs superior, it performs very poorly when applied to the real market data. On the other hand, shrinking towards w_{ew} improves the performance substantially both before and after transaction costs.

The difference between the reference portfolios becomes more evident in utility maximization: shrinking towards w_{ew} considerably improves the performance of all models, whereas using w_{t-} hardly adds value. When a portfolio is subject to large estimation errors, the current portfolio can be substantially different from the true optimal portfolio, and penalizing the deviation from it does not necessarily improve performance even after transaction costs.

As shown in Table X, the performance of most portfolios deteriorates without the variance constraint and outperforming EW becomes more challenging. This is because utility maximizing portfolios are more sensitive to the mean. In fact, EW performs considerably worse if its weights are adjusted so as to maximize utility (unreported). The performance of the optimal portfolios that ignore parameter uncertainty, *e.g.*, ML and OC, drops most when the variance constraint is removed. KZ, TZML, and TZKZ without the deviation penalty perform relatively better, but their performances also deteriorate substantially. When augmented with the deviation penalty, however, these models perform robustly and their normalized CE's are not particularly lower than those from variance targeting. It is also worth noting that a higher degree of penalization is required to maximize performance in utility maximization.

Table XI reports calibrated δ values. As expected, a larger δ is required when estimation errors have a greater impact on portfolio performance: *i.e.*, when the window size is small, transaction costs are taken into account, and the optimization criterion is utility maximization. In an unreported analysis, the models with the calibrated δ are compared with models with a constant δ (1, 2, or 3) and found to outperform all these models, confirming the effectiveness of the proposed calibration method.

[TABLE XI]

5.3 Detailed Analysis of Empirical Results

This section provides detailed analyses of the empirical results reported in Table IX and X.

Market Portfolio Apart from the equal-weight portfolio, the market portfolio (MKT) and the S&P500 index (SP500) are also used as benchmarks. The market portfolio for the first dataset is proxied by the MSCI world index including the US, and the market portfolio for the other datasets is proxied by the CRSP value-weighted market portfolio.

As can be seen from Table X, MKT and SP500 perform comparably to each other and to EW: their normalized mean CE's are respectively 0.22 and 0.24, whereas that of EW is 0.24, and they outperform EW four and five times respectively. Their relative performance improves slightly after accounting for transaction costs as they do not incur transaction costs. MKT and SP500 appear to outperform EW in Table IX, but this is because all portfolios except MKT and SP500 are adjusted so as to satisfy the variance target.

The deviation penalty models show superior performance when compared to MKT or SP500, and the performance difference is similar to the difference from EW.

ML vs. TZML vs. DPML Under variance targeting, ML outperforms EW in ten datasets and has a considerably higher mean CE before transaction costs. While much of the advantage disappears after transaction costs, it still outperforms EW in nine datasets. The benefit of combining ML with EW is evident from TZML and DPML. The mean CE's of TZML(δ^*) and DPML(δ^*) after transaction costs are respectively 0.558 and 0.553, whereas those of ML and EW are respectively 0.135 and 0.246, and they outperform EW in all datasets before transaction costs and in eleven datasets after transaction costs, when $T = 120$ and under variance targeting. The improvement over ML is more noticeable when T is small.

KZ vs. TZML vs. TZKZ Consistent with the findings of Tu and Zhou (2011), TZKZ outperforms TZML and KZ. This can be anticipated to some extent as TZKZ involves three portfolios, whereas the others involve only two. Between KZ and TZML, TZML yields higher CE's, whilst KZ outperforms EW more often. Overall, both KZ and TZML perform well when T is large, but the performance of KZ deteriorates rapidly as T decreases. This can be attributed to the fact that both ML and MV which comprise KZ depend on the input parameters and therefore subject to estimation errors. As discussed below, the performance of KZ improves markedly when it is augmented with the deviation penalty.

Effects of Deviation Penalty Penalizing the deviation from w_{ew} improves portfolio performance in most cases regardless of the base model. Improvements are particularly noticeable in the presence of transaction costs and in terms of the outperformance ratio. Among KZ, TZML, TZKZ, and DPML, KZ benefits most by incorporating the deviation penalty. This is because the original KZ, contrary to the others, does not involve w_{ew}. Comparing KZ(δ^*) with TZKZ(δ^*), KZ(δ^*) tends to perform marginally better even though both models involve the same three portfolios.

The models with $w_0 = w_{t-}$, *i.e.*, KZc, TZMLc, TZKZc, and DPMLc, perform rather disappointingly. When these are compared with their counterparts with $w_0 = w_{ew}$, the latter models almost always perform better with respect to all criteria. The poor performance of DPMLc is particularly noticeable, which indicates that the estimation errors in the actual market data are substantially higher than assumed: in simulations, DPMLc outperformed DPML when estimation errors were small ($T = 120$ or 240) but was outperformed otherwise (T=60).

The above finding conveys an important message as the deviation penalty models with $w_0 = w_{t-}$ resemble the models that take transaction costs into account (*e.g.*, Gârleanu and Pedersen, 2013; DeMiguel et al., 2015; Olivares-Nadal and DeMiguel, 2018). Shrinking towards the current portfolio does improve portfolio performance but is less effective than shrinking towards the equal-weight portfolio regardless of transaction costs. It appears that a certain degree of robustness should be assured beforehand in order to benefit from the former approach.

Estimation of K In general, DPMLK slightly outperforms DPML and requires lower δ values. When δ is set to 0, the difference between DPMLK and DPML becomes more prominent (unreported). This suggests that the proposed estimation method for K addresses the uncertainty in mean more adequately than the simple assumption of $K = T$.

Effects of Short-sale Constraint When optimal portfolio models are subject to the short-sale constraint, they perform robustly and outperform EW more frequently compared with their unconstrained counterparts. In fact, the short-sale constrained deviation penalty models (DPML+ and DPMV+) are among the best performers in terms of the outperformance ratio and demonstrate robust performance under utility maximization. They also have a significantly lower turnover and leverage. VT also performs robustly. This is because VT is implicitly short-sale constrained and depends only on the cross-sectional variation of the variances which is stable over time. Nevertheless, these models' performances are rather suppressed as evidenced by the low CE's. The short-sale constraint is an effective tool to enhance robustness especially when parameter uncertainty is large, but at the same time, it hinders high return potential.

Effects of Estimation Window Size Asset returns are not stationary over a long period and using a large estimation window does not necessarily lead to smaller estimation errors. Determining the optimal estimation window size mainly depends on two aspects: estimation errors and transaction costs.

Based on the performance before transaction costs, many portfolio models turn out to perform best when $T = 120$ and worst when $T = 60$. It appears that estimation errors decline with T until some point and then increase again. The window size also has an effect on the portfolio loadings of the shrinkage estimators: a larger T will put more weight on $\hat{w}_{ml}$ regardless of the actual estimation errors, and their performance could deteriorate rapidly beyond a certain T. On the other hand, a larger window size is always beneficial in terms of transaction costs as the moment estimates become more persistent resulting in lower turnover.

The window size should be determined considering several factors such as portfolio strategy, actual transaction costs, and dataset. Nevertheless, for the datasets considered in this paper, $T = 120$ seems to be a reasonable choice, especially for the deviation penalty models. Although not pursued in this paper, applying different window sizes to mean and covariance estimations may improve overall estimation accuracy.

5.4 Robustness Check

Comprehensive robustness tests are conducted to verify the earlier findings. This includes sub-period analyses and tests on ten additional datasets which do not contain the market and factor portfolios, and other robustness checks.

Sub-Period Performance Figure 4 reports the sub-period performance of the deviation penalty models. It reveals that the deviation penalty models perform consistently over time. All four models maintain their superior performance against EW across sub-periods.

[FIGURE 4]

Economic Cycle To examine whether the deviation penalty models maintain their superior performance in different economic cycles, they are tested during recession periods and the remaining periods. The NBER recession indicator is used to identify recession months during the sample period.

Table XII compares the performance of the deviation penalty models from these two periods. As dataset D1 has a shorter sample period, the results are obtained from datasets D2-D13, and therefore the maximum number of outperformance is 12 instead of 13.

The proposed models perform consistently well in both samples outperforming EW in most datasets, especially when the estimation window size is large. One might expect a shorter window size more appropriate during recession periods as it would reflect recent economic changes better. However, the results suggest that a long window size (at least 120 months in the test) is still better regardless of the economic cycle.

In an unreported analysis, the portfolio models are also examined over the market crash in October 1987, the dot-com bubble burst from 2000 to 2002, and the global financial crisis in 2008. While the deviation penalty models have negative returns in October 1987 and over the financial crisis, they still outperform the market and the equal-weight portfolio in all three periods.

[TABLE XII]

Different Rebalancing Frequencies The portfolio models are examined assuming either quarterly or annual rebalancing, and the results are reported in the Internet Appendix. When portfolios are rebalanced less frequently, there is a tradeoff between increased deviation from the optimal portfolio and reduced transaction costs. As expected, when the deviation penalty models are rebalanced less frequently, they tend to perform worse before transaction costs, but still outperform EW in most datasets. The performance after transaction costs improves slightly when the portfolios are rebalanced quarterly but it becomes worse when rebalanced annually, suggesting that the performance deterioration due to deviation from the optimal portfolio outweighs the gain from reduced transaction costs. It appears that quarterly is a good rebalancing frequency for the portfolios that allow short-sale. For the portfolios with short-sale constraints, more frequent rebalancing appears to work better as these portfolios do not incur much transaction cost.

Assets Excluding Factor Portfolios When the models are tested on the new datasets excluding factor portfolios, the results are qualitatively similar to those presented here, but the resulting portfolios are usually less leveraged (see the Internet Appendix). This is because it is no longer possible to short the market and buy other assets.

Weight Constraints In an unreported analysis, an additional optimization criterion, variance targeting with weight constraints, $|w_i| \leq 0.5$, is also tested. The purpose of this criterion is to generate more realistic portfolios with low leverage. Imposing the weight constraints does not alter the results considerably, and the rankings of the models are largely preserved. The CE and SR tend to be slightly smaller with the constraints, but the optimal portfolios perform more consistently across the datasets and outperform EW more frequently. The overall effect of the weight constraint is similar to that of the short-sale constraint, but, with a relaxed lower bound, it appears to strike a better balance between robustness and performance.

By and large, the conclusions drawn in this paper remain valid in the additional datasets and optimization criterion.

The empirical results favour the equal-weight portfolio as the reference portfolio. A natural question that arises would then be whether any fixed-weight portfolio could yield the same performance. While this topic is not pursued further, there are a few reasons to favour the equal-weight portfolio over other fixed-weight portfolios. First of all, it is economically meaningful as it assumes that all the assets have the same return-risk ratio, which is in line with the capital asset pricing model. Secondly, if we randomly choose a portfolio under (strict) short-sale constraints, the expected portfolio is the equal-weight portfolio. Furthermore, if the number of assets increases, it converges to the equal-weight portfolio (See the Internet Appendix for proof).

6 Concluding Remarks

This paper develops a new portfolio choice model that penalizes the deviation from a reference portfolio. Penalizing the deviation renders a biased, but robust portfolio that performs superior under parameter uncertainty.

Existing shrinkage portfolio models can be improved by incorporating the deviation penalty. These models are sub-optimal as they only address input parameter uncertainty and fail to recognize the uncertainty inherited to model parameters. They are also vulnerable to misspecification. The added robustness by the deviation penalty mitigates these problems and enhances performance considerably.

Two reference portfolios, the current portfolio and the equal-weight portfolio, are considered, and it turns out that the latter is a much better choice than the former even in the presence of transaction costs. Indeed, the latter incurs lower transaction costs. This result is contrary to the widely-accepted belief that accounting for transaction costs improves portfolio performance and reduces transaction costs: this must be true, but the effect appears rather trivial compared to using the equal-weight portfolio.

A data-driven calibration method to determine the degree of penalization is offered. This method is readily implementable and revealed to be effective, generating superior performance when applied to various models and circumstances. Comprehensive simulation and empirical studies involving different sample periods, datasets, and optimization criteria support our model and confirm above claims.

A Utility Maximization

A.1 Proof of Proposition 1

The expected utility can be rearranged as follows:

$$\begin{aligned}E[U(a,b)] = aE[\hat{w}_{ml}]'\mu + bw_0'\mu - \frac{\gamma}{2}\left(a^2E[\hat{w}_{ml}'\Sigma\hat{w}_{ml}] + 2abE[\hat{w}_{ml}]'\Sigma w_0 + b^2w_0'\Sigma w_0\right) \\ - \frac{\delta}{2}\left(a^2E[\hat{w}_{ml}'\Sigma\hat{w}_{ml}] + 2a(b-1)E[\hat{w}_{ml}]'\Sigma w_0 + (b-1)^2w_0'\Sigma w_0\right).\end{aligned} \tag{A.1}$$

Differentiating the expected utility with respect to a and b, the first order conditions are given by

$$\begin{aligned}\frac{\partial E[U(a,b)]}{\partial a} &= E[\hat{w}_{ml}]'\mu - a\gamma E[\hat{w}_{ml}'\Sigma\hat{w}_{ml}] - b\gamma E[\hat{w}_{ml}]'\Sigma w_0 \\ &\quad - a\delta E[\hat{w}_{ml}'\Sigma\hat{w}_{ml}] - (b-1)\delta E[\hat{w}_{ml}]'\Sigma w_0 = 0, \qquad (A.2) \\ \frac{\partial E[U(a,b)]}{\partial b} &= w_0'\mu - a\gamma E[\hat{w}_{ml}]'\Sigma w_0 - b\gamma w_0'\Sigma w_0 \\ &\quad - a\delta E[\hat{w}_{ml}]'\Sigma w_0 - (b-1)\delta w_0'\Sigma w_0 = 0. \qquad (A.3)\end{aligned}$$

Solving for a and b, the optimal parameters are obtained:

$$a^* = \frac{1}{(\gamma+\delta)}\frac{E[\hat{w}_{ml}]'\mu - w_0'\mu\dfrac{E[\hat{w}_{ml}]'\Sigma w_0}{w_0'\Sigma w_0}}{E[\hat{w}_{ml}'\Sigma\hat{w}_{ml}] - E[\hat{w}_{ml}]'\Sigma w_0\dfrac{E[\hat{w}_{ml}]'\Sigma w_0}{w_0'\Sigma w_0}}, \tag{A.4}$$

$$b^* = \frac{1}{(\gamma+\delta)}\frac{E[\hat{w}_{ml}]'\mu - w_0'\mu\dfrac{E[\hat{w}_{ml}'\Sigma\hat{w}_{ml}]}{E[\hat{w}_{ml}]'\Sigma w_0}}{E[\hat{w}_{ml}]'\Sigma w_0 - w_0'\Sigma w_0\dfrac{E[\hat{w}_{ml}'\Sigma\hat{w}_{ml}]}{E[\hat{w}_{ml}]'\Sigma w_0}} + \frac{\delta}{\gamma+\delta}. \tag{A.5}$$

Since $\hat{\mu}$ and $\hat{\Sigma}$ are independent of each other and

$$\hat{\mu} \sim N\left(\mu, \frac{\Sigma}{K}\right), \quad \hat{\Sigma} \sim \mathcal{W}_N(T-1,\Sigma)\frac{1}{T}, \tag{A.6}$$

it follows that

$$E[\tilde{\Sigma}^{-1}] = \Sigma^{-1}, \quad E[\hat{\mu}\Sigma^{-1}\hat{\mu}] = \frac{N}{K} + \theta^2, \tag{A.7}$$

where $\theta^2 = \mu'\Sigma^{-1}\mu$. The latter equation is from $K\hat{\mu}\Sigma^{-1}\hat{\mu} \sim \chi^2_N(K\mu'\Sigma^{-1}\mu)$. Also, it can be shown that (see Kan and Zhou (2007) and the references therein)

$$E[\hat{\mu}'\tilde{\Sigma}^{-1}\Sigma\tilde{\Sigma}^{-1}\hat{\mu}] = c_1\left(\frac{N}{K} + \theta^2\right), \tag{A.8}$$

where $c_1 = \frac{(T-2)(T-N-2)}{(T-N-1)(T-N-4)}$. Then,

$$E[\hat{w}_{ml}] = \frac{1}{\gamma}\Sigma^{-1}\mu, \tag{A.9}$$

$$E[\hat{w}_{ml}'\Sigma\hat{w}_{ml}] = \frac{c_1}{\gamma^2}\left(\frac{N}{K} + \theta^2\right). \tag{A.10}$$

Substituting (A.9) and (A.10) into (A.4) and (A.5), the optimal parameters are rewritten as follows:

$$a^* = \frac{\gamma}{\gamma+\delta}a_0^*, \tag{A.11}$$

$$b^* = \frac{\gamma}{\gamma+\delta}b_0^* + \frac{\delta}{\gamma+\delta}, \tag{A.12}$$

where

$$a_0^* = \frac{\theta^2 - \psi^2}{c_1\left(\frac{N}{K}+\theta^2\right) - \psi^2}, \tag{A.13}$$

$$b_0^* = \frac{c_1\left(\frac{N}{K}+\theta^2\right) - \theta^2}{c_1\left(\frac{N}{K}+\theta^2\right) - \psi^2}\frac{1}{\gamma}\frac{w_0'\mu}{w_0'\Sigma w_0}, \tag{A.14}$$

$$\psi^2 = \mu_0'\Sigma^{-1}\mu, \quad \mu_0 = \frac{w_0'\mu}{w_0'\Sigma w_0}\Sigma w_0. \tag{A.15}$$

A.2 Estimation of a^* and b^*

Estimation of a^* and b^* involves estimation of θ^2, ψ^2, and w_{im}. For θ^2 and ψ^2, the method proposed by Kan and Zhou (2007) is adopted with modification for the different distributional assumption of $\hat{\mu}$.

- Estimation of θ^2

  Since

$$K\hat{\mu}\Sigma^{-1}\hat{\mu} \sim \chi^2_N(K\mu'\Sigma^{-1}\mu), \quad \frac{\hat{\mu}'\Sigma^{-1}\hat{\mu}}{\hat{\mu}'(T\hat{\Sigma})^{-1}\hat{\mu}} \sim \chi^2_{T-N}, \tag{A.16}$$

  and they are independent of each other, it follows that

$$\frac{K}{T}\hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu} \sim \frac{N}{T-N}\mathcal{F}_{N,T-N}(K\mu'\Sigma^{-1}\mu), \tag{A.17}$$

  where $\mathcal{F}$ is a noncentral F-distribution. Following the proof in the appendix of Kan and Zhou (2007), the estimate of θ^2 is given by

$$\tilde{\theta}^2 = \frac{(T-N-2)\hat{\theta}^2 - N}{K} + \frac{2(\hat{\theta}^2)^{N/2}(1+\hat{\theta}^2)^{-(T-2)/2}}{KB_{\hat{\theta}^2/(1+\hat{\theta}^2)}(N/2,(T-N)/2)}, \tag{A.18}$$

  where

$$\hat{\theta}^2 = \frac{K}{T}\hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}, \tag{A.19}$$

  and $B_x(a,b) = \int_0^x y^{a-1}(1-y)^{b-1}dy$ is an incomplete beta function.

- Estimation of ψ^2

  Since

$$\frac{K(w_0'\hat{\mu})^2}{w_0'\Sigma w_0} \sim \chi^2_1\left(\frac{K(w_0'\mu)^2}{w_0'\Sigma w_0}\right), \quad \frac{Tw_0'\hat{\Sigma}w_0}{w_0'\Sigma w_0} \sim \chi^2_{T-1}, \tag{A.20}$$

  and they are independent of each other, it follows that

$$\frac{K}{T}\frac{K(w_0'\hat{\mu})^2}{w_0'\hat{\Sigma}w_0} \sim \frac{1}{T-1}\mathcal{F}_{1,T-1}\left(\frac{K(w_0'\mu)^2}{w_0'\Sigma w_0}\right). \tag{A.21}$$

The estimate of ψ^2 is then given by

$$\tilde{\psi}^2 = \frac{(T-3)\hat{\psi}^2 - 1}{K} + \frac{2(\hat{\psi}^2)^{1/2}(1+\hat{\psi}^2)^{-(T-2)/2}}{KB_{\hat{\psi}^2/(1+\hat{\psi}^2)}(1/2,(T-1)/2)}, \tag{A.22}$$

where

$$\hat{\psi}^2 = \frac{K}{T}\frac{(w_0'\hat{\mu})^2}{w_0'\hat{\Sigma}w_0}. \tag{A.23}$$

- Estimation of w_{im}

  From $Tw_0'\hat{\Sigma}w_0 \sim w_0'\Sigma w_0 \cdot \chi^2_{T-1}$,

$$\frac{w_0'\Sigma w_0}{Tw_0'\hat{\Sigma}w_0} \sim \text{inv-}\chi^2_{T-1}, \tag{A.24}$$

  and

$$E\left[\frac{1}{w_0'\hat{\Sigma}w_0}\right] = \frac{T}{T-3}\frac{1}{w_0'\Sigma w_0}. \tag{A.25}$$

  Therefore, an unbiased estimate of w_{im} is given by

$$\tilde{w}_{im} = \frac{1}{\gamma}\frac{T-3}{T}\frac{w_0'\hat{\mu}}{w_0'\hat{\Sigma}w_0}w_0. \tag{A.26}$$

Simulation studies suggest that the optimal portfolio with the above estimates sometimes underperforms the more restricted model of Tu and Zhou (2011), especially when T is small. Meanwhile, assuming $w_{im} = \frac{c}{\gamma}w_0$ for some constant c appears to yield more robust performance. This can be justified as $\frac{w_0'\mu}{w_0'\Sigma w_0}$ should be constant when the returns are $i.i.d.$ Accordingly, $w_{im} = \frac{c}{\gamma}w_0$ with $c = 3$ rather than $\tilde{w}_{im}$ is used in the empirical studies of the paper.

A.3 Estimation of K

Let T_c denote a month during the sample period. For the first 119 months into the sample period, $K = T$ is assumed. When $T_c \geq 120$, the covariance matrix of $\hat{\mu}$, $\Sigma_{\hat{\mu}}$, is estimated employing the method of Lo and MacKinlay (1988):

$$\tilde{\Sigma}_{\hat{\mu}} = \frac{T_c + T - 1}{T_c - 1}\hat{\Sigma}_{\hat{\mu}}, \tag{A.27}$$

where $\hat{\Sigma}_{\hat{\mu}}$ is the ML estimate of $\Sigma_{\hat{\mu}}$. Let $\bar{\Sigma}$ denote the average of $\hat{\Sigma}$:

$$\bar{\Sigma} = \frac{1}{T_c}\sum_{t=1}^{T_c}\hat{\Sigma}_t, \tag{A.28}$$

where $\hat{\Sigma}_t$ is $\hat{\Sigma}$ at month t. K is determined so that the distance between $\frac{1}{K}\bar{\Sigma}$ and $\tilde{\Sigma}_{\hat{\mu}}$ is minimized. Defining the distance as the Frobenius norm of the lower triangular part of $(\bar{\Sigma} - K\tilde{\Sigma}_{\hat{\mu}})$, K is obtained from

$$K = \frac{v_1'v_2}{v_1'v_1}, \quad v_1 = vech(\tilde{\Sigma}_{\hat{\mu}}),\ v_2 = vech(\bar{\Sigma}), \tag{A.29}$$

where $vech(\cdot)$ is the half-vectorization operator.

B Variance Minimization

B.1 Proof of Proposition 2

Differentiating the expected variance with respect to a, the first order condition is given by

$$\frac{\partial E[V(a)]}{\partial a} = (1+\delta)aE[(\hat{w}_{mv} - w_0)'\Sigma(\hat{w}_{mv} - w_0)] + E[(\hat{w}_{mv} - w_0)'\Sigma w_0] = 0. \tag{B.1}$$

Solving for a,

$$a^* = \frac{1}{1+\delta}\frac{w_0'\Sigma w_0 - E[\hat{w}_{mv}]'\Sigma w_0}{E[\hat{w}_{mv}'\Sigma\hat{w}_{mv}] + w_0'\Sigma w_0 - 2E[\hat{w}_{mv}]'\Sigma w_0}. \tag{B.2}$$

From Kan and Smith (2008),

$$E[\hat{w}_{mv}] = \frac{\Sigma^{-1}1_N}{1_N'\Sigma^{-1}1_N}, \tag{B.3}$$

$$E[\hat{w}_{mv}'\Sigma\hat{w}_{mv}] = \frac{T-2}{T-N-1}\frac{1}{1_N'\Sigma^{-1}1_N}. \tag{B.4}$$

Substituting (B.3) and (B.4) into (B.2),

$$a^* = \frac{1}{1+\delta}\frac{\sigma_0^2 - \sigma_{mv}^2}{\sigma_0^2 - \left(1 - \frac{N-1}{T-N-1}\right)\sigma_{mv}^2}, \tag{B.5}$$

where $\sigma_0^2 = w_0'\Sigma w_0$ and $\sigma_{mv}^2 = w_{mv}'\Sigma w_{mv} = (1_N'\Sigma^{-1}1_N)^{-1}$ are the variances of w_0 and w_{mv}, respectively.

B.2 Estimation of a^*

Unbiased estimates of σ_0^2 and σ_{mv}^2 can be obtained as follows:

$$\tilde{\sigma}_0^2 = \frac{T}{T-1}w_0'\hat{\Sigma}w_0, \quad \tilde{\sigma}_{mv}^2 = \frac{T}{T-N}\frac{1}{1_N'\hat{\Sigma}^{-1}1_N}. \tag{B.6}$$

It can be seen that with the above estimates, $0 < a^* < 1$.

C Moments of Optimal Portfolio Weights

C.1 Proof of Proposition 3

When $w_0 = w_{t-}$, the deviation penalty portfolio at time t, w_t^c, can be written as

$$\begin{aligned}
w_t^c &= (1-\alpha)w_t^* + \alpha w_{t-1}^c \\
&= (1-\alpha)w_t^* + \alpha((1-\alpha)w_{t-1}^* + \alpha w_{t-2}^c) \\
&\quad\vdots \\
&= (1-\alpha)(w_t^* + \alpha w_{t-1}^* + \cdots + \alpha^{t-1}w_1^*) + \alpha^t w.
\end{aligned} \tag{C.1}$$

Since the returns are assumed to be *i.i.d.*, $E(w_t^*) = E(w_{t-1}^*) = \cdots = E(w_1^*)$, and it follows that

$$E(w_t^c) = (1-\alpha^t)E(w_t^*) + \alpha^t w, \tag{C.2}$$

$$V(w_{it}^c) = (1-\alpha)^2 V(w_{it}^* + \alpha w_{it-1}^* + \cdots + \alpha^{t-1}w_{i1}^*), \tag{C.3}$$

where w_{it} denotes the i-th element of w_t.

When $w_0 = w_{ew}$, the deviation penalty portfolio at time t, w_t^e, has the form

$$w_t^e = (1-\alpha)w_t^* + \alpha w_{ew}, \tag{C.4}$$

and its moments are given by

$$E(w_t^e) = (1-\alpha)E(w_t^*) + \alpha w_{ew}, \tag{C.5}$$

$$V(w_{it}^e) = (1-\alpha)^2 V(w_{it}^*). \tag{C.6}$$

As $0 < Cov(w^*_{it}, w^*_{it-1}) < 1$ for $t > 1$,

$$(1-\alpha)^2(1+\alpha^2+\cdots+\alpha^{2(t-1)})V(w^*_{it}) < V(w^c_{it}) < (1-\alpha^t)^2 V(w^*_{it}). \tag{C.7}$$

Therefore,

$$V(w^e_{it}) < V(w^c_{it}) < V(w^*_{it}). \tag{C.8}$$

C.2 Proof of Proposition 4

Note that for $t > 1$,

$$\Delta w^c_{it} = (1-\alpha)\Delta w^*_{it} + \alpha \Delta w^c_{it-1}, \tag{C.9}$$

$$\Delta w^e_{it} = (1-\alpha)\Delta w^*_{it}. \tag{C.10}$$

Therefore,

$$E\left[(\Delta w^c_{it})^2\right] = (1-\alpha)^2 E\left[(\Delta w^*_{it})^2\right] + \alpha^2 E\left[(\Delta w^c_{it-1})^2\right] + 2\alpha(1-\alpha)E\left[\Delta w^*_{it}\Delta w^c_{it-1}\right], \tag{C.11}$$

$$E\left[(\Delta w^e_{it})^2\right] = (1-\alpha)^2 E\left[(\Delta w^*_{it})^2\right]. \tag{C.12}$$

If $E\left[\Delta w^*_{it}\Delta w^c_{it-1}\right] > -\frac{\alpha}{2(1-\alpha)} E\left[(\Delta w^c_{it-1})^2\right]$,

$$E\left[(\Delta w^e_{it})^2\right] < E\left[(\Delta w^c_{it})^2\right]. \tag{C.13}$$

Since $\Delta w^c_{i1} = (1-\alpha)(w^*_{i1} - w_i)$ and $\Delta w^*_{i1} = (w^*_{i1} - w_i)$, $E[(\Delta w^c_{i1})^2] < E[(\Delta w^*_{i1})^2]$. From (C.9), it follows that

$$E\left[(\Delta w^c_{it})^2\right] < E\left[(\Delta w^*_{it})^2\right]. \tag{C.14}$$

Proof of Corollary 1 follows immediately from $E\left[(\Delta w^c_{it-1})^2\right] < E[(\Delta w^c_{it})^2]$.

References

Avramov, D. and Zhou, G. (2010). Bayesian portfolio analysis. *The Annual Review of Financial Economics*, 2(1):25–47.

Bessler, W., Opfer, H., and Wolff, D. (2014). Multi-asset portfolio optimization and out-of-sample performance: an evaluation of Black–Litterman, mean-variance, and naïve diversification approaches. *The European Journal of Finance*, pages 1–30.

Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43.

Brown, S. J. (1976). *Optimal portfolio choice under uncertainty: A Bayesian approach*. PhD thesis, University of Chicago, Graduate School of Business.

Brown, S. J. (1978). The portfolio choice problem: comparison of certainty equivalence and optimal bayes portfolios. *Communications in Statistics-Simulation and Computation*, 7(4):321–334.

Cao, H. H., Han, B., Hirshleifer, D., and Zhang, H. H. (2009). Fear of the unknown: Familiarity and economic decisions. *Review of Finance*, 15(1):173–206.

Ceria, S. and Stubbs, R. A. (2016). Incorporating estimation errors into portfolio selection: Robust portfolio construction. In *Asset Management*, pages 270–294. Springer.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, 22(5):1915–1953.

DeMiguel, V., Martín-Utrera, A., and Nogales, F. J. (2015). Parameter uncertainty in multiperiod portfolio optimization with transaction costs. *Journal of Financial and Quantitative Analysis*, 50(06):1443–1471.

Fabozzi, F. J., Kolm, P. N., Pachamanova, D. A., and Focardi, S. M. (2007). *Robust portfolio optimization and management*. John Wiley & Sons.

Gârleanu, N. and Pedersen, L. H. (2013). Dynamic trading with predictable returns and transaction costs. *The Journal of Finance*, 68(6):2309–2340.

Goldfarb, D. and Iyengar, G. (2003). Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1684.

Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *The Journal of Financial and Quantitative Analysis*, 21(3):279–292.

Kan, R. and Smith, D. R. (2008). The distribution of the sample minimum-variance frontier. *Management Science*, 54(7):1364–1380.

Kan, R., Wang, X., and Zhou, G. (2016). On the value of portfolio optimization in the presence of estimation risk: the case with and without risk-free asset. *Working Paper*.

Kan, R. and Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(03):621–656.

Kirby, C. and Ostdiek, B. (2012). It's all in the timing: Simple active portfolio strategies that outperform naïve diversification. *The Journal of Financial and Quantitative Analysis*, 47(2):437–467.

Klein, R. W. and Bawa, V. S. (1976). The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, 3(3):215–231.

Ledoit, O. and Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119.

Lo, A. W. and MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of financial studies*, 1(1):41–66.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Olivares-Nadal, A. V. and DeMiguel, V. (2018). A robust perspective on transaction costs in portfolio optimization. *Operations Research*.

Pástor, L. (2000). Portfolio selection and asset pricing models. *The Journal of Finance*, 55(1):179–223.

Pástor, L. and Stambaugh, R. F. (2000). Comparing asset pricing models: an investment perspective. *Journal of Financial Economics*, 56(3):335–381.

Tu, J. and Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, 99(1):204–215.

Table I: The Datasets

This table lists the datasets used in the simulation and empirical studies. The 8 international indices in D1 are the gross returns on large/mid-cap stocks from eight countries: Canada, France, Germany, Italy, Japan, Switzerland, United Kingdom, and USA. The 20 portfolios with size-sort (D5, 6, 7, 11, 12, 20) are from the corresponding 25 portfolios excluding the 5 largest portfolios. All data are from K. French website (`http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`) except D1, which is from the MSCI website (`https://www.msci.com/end-of-day-data-country`).

Dataset	Description	N	Sample Period
D1	8 International + World Indices	9	1990.10 - 2015.12
D2	10 Industry Portfolios + Market	11	1951.01 - 2015.12
D3	30 Industry Portfolios + Market	31	1951.01 - 2015.12
D4	3 Fama-French (FF) Factors	3	1951.01 - 2015.12
D5	20 FF Portfolios + Market	21	1951.01 - 2015.12
D6	20 FF Portfolios + FF 3	23	1951.01 - 2015.12
D7	20 FF Portfolios + FF 3 and Momentum	24	1951.01 - 2015.12
D8	10 Momentum Portfolios + Market	11	1951.01 - 2015.12
D9	10 Short-Term Reversal Portfolios + Market	11	1951.01 - 2015.12
D10	10 Long-Term Reversal Portfolios + Market	11	1951.01 - 2015.12
D11	20 Size/Momentum Portfolios + Market	21	1951.01 - 2015.12
D12	20 Size/Short-Term Reversal Portfolios + Market	21	1951.01 - 2015.12
D13	20 Size/Long-Term Reversal Portfolios + Market	21	1951.01 - 2015.12

Table II: *Ex-post* Optimal Portfolio Weights

This table summarizes the *ex-post* optimal portfolio weights from each dataset. The *ex-post* optimal portfolio is defined as the Markowitz portfolio obtained from the sample moments of the entire sample. '$\min w_i$' and '$\max w_i$' are respectively the minimum and maximum weights on the risky assets, '$\sum w_i$' is the sum of the risky asset weights, *i.e.*, the weight of the risky portfolio, and '$\sum |w_i|$', the sum of the absolute values of the weights, measures the degree of leverage. The last three columns are the frequency of negative expected returns on the global minimum-variance portfolio during the sample period for the estimation window size $T = 60$, 120, and 240.

	$\min w_i$	$\max w_i$	$\sum w_i$	$\sum \lvert w_i \rvert$	$P(\mu_g < 0)$		
					60	120	240
D1	-5.37	3.87	1.40	12.20	0.28	0.13	0.00
D2	-6.91	1.79	1.89	15.71	0.09	0.02	0.00
D3	-7.03	1.55	1.92	21.43	0.23	0.13	0.00
D4	0.53	2.12	3.98	3.98	0.10	0.02	0.00
D5	-2.32	3.54	2.34	26.74	0.22	0.09	0.00
D6	-4.17	3.89	-4.19	38.56	0.63	0.74	0.87
D7	-5.75	3.43	-3.64	41.90	0.66	0.75	0.88
D8	-4.02	2.47	1.53	14.40	0.17	0.11	0.06
D9	-1.42	1.45	1.51	9.64	0.23	0.19	0.12
D10	-3.55	1.31	1.46	10.67	0.15	0.08	0.00
D11	-4.23	3.79	2.39	25.21	0.11	0.00	0.00
D12	-7.12	3.65	1.24	31.71	0.20	0.22	0.15
D13	-2.65	2.37	2.21	18.79	0.14	0.01	0.00

Table III: The Portfolio Models

This table lists the portfolio models considered in the simulation and empirical studies. The models with '+' in their abbreviation are those subject to the short-sale constraint. The short-sale constraint is applied only to risky assets. Implementation details of each model can be found in the Internet Appendix.

Abbreviation	Description
W*	*Ex-post* optimal portfolio, *i.e.*, the Markowitz portfolio obtained from the sample moments of the entire sample.
EW	Equal-weight portfolio.
Classical Portfolio Strategies	
ML, ML+	Markowitz (1952) mean-variance portfolio.
MV, MV+	Global minimum-variance portfolio.
Kirby and Ostdiek (2012)	
VT	Volatility timing strategy.
OC, OC+	Optimal constrained portfolio: the Markowitz portfolio without the risk-free asset.
Kan and Zhou (2007)	
KZ	Kan and Zhou (2007) three-fund rule.
KZ(δ), KZc(δ)	KZ with deviation penalty. KZ(δ): $w_0 = w_{ew}$; KZc(δ): $w_0 = w_{t-}$.
Tu and Zhou (2011)	
TZML	Tu and Zhou (2011) model that combines ML with EW.
TZML(δ), TZMLc(δ)	TZML with deviation penalty. TZML(δ): $w_0 = w_{ew}$; TZMLc(δ): $w_0 = w_{t-}$.
TZKZ	Tu and Zhou (2011) model that combines KZ with EW.
TZKZ(δ), TZKZc(δ)	TZKZ with deviation penalty. TZKZ(δ): $w_0 = w_{ew}$; TZKZc(δ): $w_0 = w_{t-}$.
Deviation Penalty Models	
DPML(δ), DPML+(δ)	Utility maximization with deviation penalty. $w_0 = w_{ew}$.
DPMLc(δ), DPMLc+(δ)	Utility maximization with deviation penalty. $w_0 = w_{t-}$.
DPMLK(δ)	DPML(δ) with estimated K.
DPMV(δ), DPMV+(δ)	Variance minimization with deviation penalty. $w_0 = w_{ew}$.
DPMVc(δ), DPMVc+(δ)	Variance minimization with deviation penalty. $w_0 = w_{t-}$.
δ: deviation penalty coefficient	

Table IV: Expected Utility: $\gamma = 3$

This table reports the mean and standard error of the utilities of selected portfolios, obtained from 10,000 iterations. Utilities are normalized by that of W*. The reported values are averages across the datasets; D1, D2, D5, and D8. The numbers in the header are estimation window sizes. EW and EW* are equal-weight portfolios adjusted to maximize utility, respectively using sample and true moments.

	Mean					Standard Error				
	60	120	240	360	600	60	120	240	360	600
W*	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000
EW*	0.230	0.230	0.230	0.230	0.230	0.000	0.000	0.000	0.000	0.000
EW	-0.024	0.110	0.172	0.191	0.207	0.407	0.183	0.082	0.056	0.032
ML	-6.094	-1.074	0.167	0.480	0.705	4.850	1.094	0.387	0.233	0.128
KZ	0.058	0.379	0.587	0.681	0.782	0.654	0.295	0.162	0.123	0.086
TZML	0.183	0.395	0.574	0.671	0.778	0.530	0.247	0.159	0.127	0.090
TZKZ	0.259	0.436	0.581	0.664	0.764	0.363	0.183	0.135	0.117	0.091
DPML(0)	0.200	0.402	0.576	0.672	0.778	0.507	0.238	0.156	0.126	0.090
DPML(1)	0.286	0.431	0.571	0.653	0.747	0.274	0.161	0.138	0.121	0.092
DPML(2)	0.312	0.424	0.541	0.611	0.693	0.185	0.140	0.132	0.117	0.092
DPML(3)	0.318	0.409	0.509	0.570	0.642	0.145	0.128	0.124	0.111	0.088
ML+	-0.144	0.153	0.291	0.335	0.374	0.744	0.297	0.130	0.087	0.050
DPML+(0)	0.280	0.311	0.347	0.366	0.389	0.079	0.068	0.055	0.046	0.034
DPML+(1)	0.281	0.311	0.345	0.365	0.387	0.065	0.059	0.050	0.043	0.031

Table V: Expected Utility: $\gamma = 3$, Error in Mean

This table reports the mean and standard error of the utilities of selected portfolios, obtained from 10,000 iterations. The misspecification of the mean is simulated by adding $0.2\text{diag}(\mu)z$ to μ in (5), where $\text{diag}(\mu)$ is a diagonal matrix with μ in its diagonal, and z is an N-dimensional standard normal random variable. Utilities are normalized by that of W*. The reported values are averages across the datasets; D1, D2, D5, and D8. The numbers in the header are estimation window sizes. EW and EW* are equal-weight portfolios adjusted to maximize utility, respectively using sample and true moments.

	Mean					Standard Error				
	60	120	240	360	600	60	120	240	360	600
W*	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000
EW*	0.230	0.230	0.230	0.230	0.230	0.000	0.000	0.000	0.000	0.000
EW	-0.027	0.108	0.171	0.191	0.206	0.415	0.181	0.086	0.056	0.034
ML	-10.210	-3.321	-1.581	-1.139	-0.831	7.851	2.522	1.521	1.276	1.151
KZ	-0.473	-0.335	-0.315	-0.325	-0.352	1.370	1.057	0.988	0.955	0.971
TZML	-0.317	-0.302	-0.325	-0.342	-0.368	1.287	1.035	0.981	0.951	0.968
TZKZ	-0.027	0.014	-0.043	-0.097	-0.180	0.902	0.767	0.814	0.829	0.892
DPML(0)	-0.293	-0.291	-0.320	-0.338	-0.366	1.266	1.029	0.979	0.950	0.968
DPML(1)	0.039	0.091	0.121	0.133	0.141	0.681	0.558	0.541	0.532	0.546
DPML(2)	0.173	0.239	0.287	0.309	0.329	0.430	0.364	0.360	0.359	0.370
DPML(3)	0.235	0.303	0.357	0.382	0.406	0.305	0.270	0.272	0.274	0.282
ML+	-0.183	0.117	0.247	0.289	0.322	0.798	0.328	0.164	0.117	0.085
DPML+(0)	0.273	0.297	0.321	0.333	0.346	0.101	0.094	0.082	0.073	0.065
DPML+(1)	0.281	0.308	0.334	0.347	0.360	0.076	0.071	0.063	0.057	0.051

Table VI: Certainty Equivalent: $\gamma = 3$

This table reports the mean and standard error of the certainty equivalents (CE) of selected portfolios, obtained from 10,000 iterations. Portfolios are assumed to be rebalanced monthly and managed for ten years. CE's are normalized by that of W*. The reported values are averages across the datasets; D1, D2, D5, and D8. The numbers in the header are estimation window sizes. EW and EW* are equal-weight portfolios adjusted to maximize utility, respectively using sample and true moments.

	Mean					Standard Error				
	60	120	240	360	600	60	120	240	360	600
W*	1.000	1.000	1.000	1.000	1.000	0.653	0.650	0.649	0.653	0.647
EW*	0.232	0.229	0.231	0.228	0.228	0.306	0.304	0.302	0.307	0.305
EW	-0.018	0.114	0.175	0.191	0.206	0.385	0.358	0.338	0.334	0.322
ML	-5.994	-1.021	0.193	0.494	0.715	1.778	1.024	0.873	0.812	0.748
KZ	0.064	0.388	0.596	0.685	0.785	0.545	0.559	0.577	0.591	0.597
TZML	0.188	0.400	0.581	0.675	0.781	0.478	0.505	0.546	0.569	0.586
TZKZ	0.261	0.439	0.586	0.666	0.765	0.449	0.469	0.495	0.515	0.533
DPML(0)	0.205	0.407	0.583	0.676	0.781	0.470	0.500	0.541	0.565	0.584
DPML(1)	0.290	0.434	0.575	0.654	0.746	0.431	0.455	0.477	0.489	0.494
DPML(2)	0.314	0.426	0.544	0.611	0.691	0.405	0.423	0.437	0.443	0.443
DPML(3)	0.320	0.410	0.511	0.570	0.640	0.387	0.400	0.409	0.413	0.411
ML+	-0.130	0.163	0.296	0.337	0.374	0.534	0.521	0.496	0.484	0.464
DPML+(0)	0.282	0.311	0.348	0.366	0.388	0.354	0.369	0.382	0.395	0.404
DPML+(1)	0.283	0.311	0.346	0.364	0.386	0.343	0.353	0.362	0.371	0.376
DPMLc(0)	0.126	0.446	0.596	0.680	0.781	0.804	0.589	0.561	0.571	0.585
DPMLc(1)	0.226	0.483	0.600	0.667	0.750	0.700	0.510	0.478	0.482	0.488
DPMLc(2)	0.272	0.490	0.585	0.639	0.707	0.650	0.475	0.440	0.441	0.442
DPMLc(3)	0.297	0.489	0.573	0.621	0.687	0.622	0.458	0.423	0.424	0.426

(a) Before Transaction Cost

	Mean					Standard Error				
	60	120	240	360	600	60	120	240	360	600
W*	1.000	1.000	1.000	1.000	1.000	0.711	0.707	0.706	0.711	0.704
EW*	0.251	0.247	0.250	0.247	0.247	0.333	0.331	0.328	0.334	0.332
EW	-0.035	0.117	0.187	0.205	0.222	0.418	0.388	0.367	0.363	0.350
ML	-8.284	-1.655	-0.053	0.343	0.631	2.186	1.138	0.948	0.882	0.812
KZ	-0.172	0.278	0.546	0.655	0.770	0.577	0.590	0.616	0.634	0.644
TZML	0.006	0.304	0.533	0.644	0.765	0.510	0.530	0.580	0.609	0.632
TZKZ	0.137	0.387	0.569	0.661	0.770	0.473	0.494	0.526	0.551	0.574
DPML(0)	0.030	0.315	0.536	0.645	0.765	0.500	0.525	0.575	0.605	0.629
DPML(1)	0.178	0.383	0.560	0.654	0.760	0.454	0.479	0.509	0.524	0.533
DPML(2)	0.236	0.395	0.542	0.623	0.715	0.428	0.448	0.467	0.477	0.479
DPML(3)	0.261	0.391	0.517	0.588	0.669	0.410	0.426	0.439	0.446	0.445
ML+	-0.185	0.154	0.309	0.357	0.400	0.578	0.564	0.539	0.525	0.504
DPML+(0)	0.288	0.326	0.370	0.392	0.417	0.383	0.399	0.415	0.429	0.439
DPML+(1)	0.291	0.326	0.369	0.390	0.415	0.372	0.383	0.393	0.403	0.409
DPMLc(0)	0.020	0.423	0.593	0.680	0.783	0.921	0.652	0.611	0.621	0.635
DPMLc(1)	0.152	0.480	0.617	0.689	0.776	0.787	0.556	0.517	0.522	0.528
DPMLc(2)	0.214	0.495	0.608	0.668	0.741	0.725	0.517	0.476	0.477	0.479
DPMLc(3)	0.248	0.498	0.599	0.652	0.722	0.691	0.497	0.457	0.459	0.462

(b) After Transaction Cost

Table VII: Expected and Sample Variances

This table reports the mean and standard error of the variances of selected portfolios, obtained from 10,000 iterations. In the second panel, portfolios are assumed to be rebalanced monthly and managed for ten years. Variances are normalized by that of the *ex-post* global minimum-variance portfolio, MV*. The reported values are averages across the datasets; D1, D2, D5, and D8. The numbers in the header are estimation window sizes.

	Mean					Standard Error				
	60	120	240	360	600	60	120	240	360	600
MV*	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000
MV	1.276	1.115	1.054	1.035	1.020	0.127	0.049	0.022	0.014	0.008
DPMV(0)	1.220	1.103	1.051	1.034	1.020	0.099	0.044	0.021	0.014	0.008
DPMV(1)	1.281	1.230	1.202	1.192	1.183	0.071	0.046	0.030	0.023	0.018

(a) Expected Variance

	Mean					Standard Error				
	60	120	240	360	600	60	120	240	360	600
MV*	1.000	1.000	1.000	1.000	1.000	0.129	0.129	0.130	0.130	0.129
MV	1.277	1.116	1.054	1.035	1.021	0.176	0.146	0.137	0.134	0.131
DPMV(0)	1.220	1.103	1.051	1.034	1.020	0.163	0.144	0.137	0.134	0.131
DPMV(1)	1.280	1.230	1.201	1.192	1.183	0.159	0.154	0.153	0.153	0.152
DPMVc(0)	1.240	1.106	1.051	1.033	1.020	0.172	0.145	0.137	0.134	0.131
DPMVc(1)	1.222	1.102	1.051	1.034	1.021	0.169	0.145	0.137	0.134	0.132

(b) Sample Variance

Table VIII: Utility Loss due to Model Parameter Uncertainty

This table compares the certainty equivalents obtained from true (True) and estimated (Estd) model parameters of the four shrinkage models, KZ, TZML, TZKZ, and DPML. The reported values are normalized utilities averaged across the datasets; D1, D2, D5, and D8. The numbers in the header are estimation window sizes.

	60		120		240	
	True	Estd	True	Estd	True	Estd
KZ	0.176	0.058	0.447	0.388	0.633	0.596
TZML	0.360	0.184	0.491	0.401	0.635	0.581
TZKZ	0.372	0.256	0.507	0.439	0.649	0.585
DPML(0)	0.371	0.203	0.498	0.406	0.640	0.581

Table IX: Certainty Equivalent: Variance Targeting

This table reports the normalized CE's from variance targeting. 'Mean CE' is the mean of the normalized CE's across the datasets in Table I, and 'N(>EW)' is the number of the datasets in which a portfolio outperforms EW. δ^* indicates the models use a calibrated δ.

	Before Cost						After Cost					
	Mean CE			N(>EW)			Mean CE			N(>EW)		
	60	120	240	60	120	240	60	120	240	60	120	240
W*	1.000	1.000	1.000	13	13	13	1.000	1.000	1.000	13	13	13
EW	0.183	0.242	0.302	0	0	0	0.170	0.246	0.319	0	0	0
MKT	0.335	0.335	0.335	11	10	7	0.361	0.361	0.361	12	10	8
SP500	0.365	0.365	0.365	12	12	9	0.394	0.394	0.394	12	12	10
ML	0.433	0.649	0.571	9	10	9	-0.891	0.135	0.301	4	6	6
ML+	0.469	0.474	0.407	12	13	10	0.419	0.460	0.407	12	13	8
MV	-0.124	0.315	0.316	4	9	8	-1.886	-0.284	0.069	1	3	5
MV+	0.229	0.321	0.390	8	9	12	0.172	0.310	0.405	5	8	11
VT	0.205	0.252	0.334	10	10	12	0.191	0.255	0.354	10	9	12
OC	0.591	0.717	0.668	10	11	9	-0.469	0.303	0.469	5	7	7
OC+	0.418	0.398	0.387	13	12	11	0.379	0.377	0.386	12	11	10
KZ	0.402	0.742	0.668	9	13	11	-1.036	0.241	0.441	2	7	6
KZ(δ^*)	0.774	0.873	0.733	13	13	13	0.132	0.597	0.592	7	12	11
KZc(δ^*)	0.422	0.743	0.669	9	13	11	-0.429	0.386	0.505	5	8	8
TZML	0.653	0.735	0.653	11	12	10	-0.329	0.346	0.463	5	9	6
TZML(δ^*)	0.771	0.797	0.691	13	13	13	0.224	0.558	0.566	7	11	10
TZMLc(δ^*)	0.663	0.736	0.655	11	12	10	0.086	0.481	0.524	8	9	8
TZKZ	0.597	0.784	0.685	12	13	13	-0.561	0.385	0.513	4	8	9
TZKZ(δ^*)	0.770	0.829	0.704	13	13	13	0.251	0.594	0.578	7	13	11
TZKZc(δ^*)	0.604	0.786	0.686	12	13	13	-0.018	0.524	0.571	7	10	10
DPML(δ^*)	0.739	0.800	0.688	13	13	13	0.199	0.553	0.561	8	11	9
DPMLc(δ^*)	0.366	0.510	0.545	7	9	8	-0.013	0.328	0.451	7	8	8
DPMLK(δ^*)	0.748	0.793	0.679	13	13	12	0.194	0.562	0.566	7	12	10
DPML+(δ^*)	0.464	0.479	0.423	13	13	12	0.431	0.470	0.432	13	13	9
DPMV(δ^*)	0.380	0.464	0.467	13	13	13	0.211	0.361	0.419	7	10	11
DPMVc(δ^*)	0.240	0.319	0.283	8	9	8	-0.461	-0.065	0.062	4	8	7
DPMV+(δ^*)	0.450	0.466	0.421	13	13	13	0.439	0.473	0.432	13	13	13

Table X: Certainty Equivalent: Utility Maximization

This table reports the normalized CE's from utility maximization. 'Mean CE' is the mean of the normalized CE's across the datasets in Table I, and 'N(>EW)' is the number of the datasets in which a portfolio outperforms EW. δ^* indicates the models use a calibrated δ.

	Before Cost						After Cost					
	Mean CE			N(>EW)			Mean CE			N(>EW)		
	60	120	240	60	120	240	60	120	240	60	120	240
W*	1.000	1.000	1.000	13	13	13	1.000	1.000	1.000	13	13	13
EW	0.240	0.240	0.240	0	0	0	0.305	0.305	0.305	0	0	0
MKT	0.222	0.222	0.222	4	4	4	0.284	0.284	0.284	4	4	4
SP500	0.244	0.244	0.244	5	5	5	0.311	0.311	0.311	5	5	5
ML	-23.657	-2.919	-0.783	0	0	0	-70.533	-11.742	-3.146	0	0	0
ML+	-0.585	0.037	0.094	0	5	4	-1.046	-0.093	0.052	0	2	3
MV	0.224	0.270	0.256	6	10	10	0.006	0.229	0.263	0	5	5
MV+	0.244	0.251	0.250	9	9	9	0.286	0.305	0.311	7	9	8
VT	0.246	0.248	0.251	12	12	12	0.314	0.317	0.320	12	12	12
OC	-18.741	-2.576	-0.564	0	0	0	-57.187	-10.148	-2.480	0	0	0
OC+	0.275	0.251	0.268	7	8	8	0.288	0.281	0.322	7	8	8
KZ	-0.310	0.349	0.234	1	9	7	-3.005	-0.855	-0.428	0	1	1
KZ(δ^*)	0.506	0.669	0.559	13	13	13	0.115	0.504	0.521	4	9	9
KZc(δ^*)	-0.272	0.350	0.233	2	9	7	-2.041	-0.611	-0.346	0	1	1
TZML	0.022	0.419	0.317	4	9	7	-2.060	-0.575	-0.226	0	1	1
TZML(δ^*)	0.506	0.627	0.548	12	12	12	0.169	0.480	0.517	5	9	9
TZMLc(δ^*)	0.053	0.420	0.320	5	9	7	-1.244	-0.319	-0.133	0	1	2
TZKZ	0.166	0.545	0.391	7	12	8	-1.385	-0.114	0.039	0	5	3
TZKZ(δ^*)	0.503	0.642	0.532	13	13	13	0.219	0.508	0.516	6	10	9
TZKZc(δ^*)	0.179	0.545	0.391	7	12	8	-0.715	0.087	0.113	0	6	4
DPML(δ^*)	0.487	0.618	0.545	12	12	12	0.161	0.475	0.515	5	9	9
DPMLc(δ^*)	-4.894	-1.411	-0.306	0	0	4	-14.786	-3.766	-1.035	0	0	0
DPMLK(δ^*)	0.496	0.615	0.531	12	12	12	0.172	0.483	0.509	4	10	9
DPML+(δ^*)	0.347	0.359	0.333	13	13	11	0.394	0.418	0.409	13	12	9
DPMV(δ^*)	0.312	0.335	0.314	10	10	10	0.295	0.345	0.357	5	10	10
DPMVc(δ^*)	0.267	0.272	0.245	9	10	8	0.284	0.308	0.279	8	8	8
DPMV+(δ^*)	0.276	0.276	0.265	10	10	10	0.332	0.340	0.333	10	10	10

Table XI: δ Calibration

This table reports the calibrated δ's for each model with and without transaction costs. The reported values are the mean across the sample period and datasets.

	Variance Targeting						Utility Maximization					
	Before Cost			After Cost			Before Cost			After Cost		
	60	120	240	60	120	240	60	120	240	60	120	240
KZ	4.49	2.49	1.83	8.08	5.58	4.91	4.55	2.66	3.20	8.49	6.21	5.95
KZc	3.84	1.71	1.82	9.23	8.68	8.55	3.84	1.08	1.42	9.07	7.86	6.69
TZML	3.19	2.32	1.80	7.76	6.11	5.05	3.75	2.11	2.82	8.48	6.23	5.76
TZMLc	2.24	1.37	1.90	9.23	9.05	8.77	3.82	0.98	1.24	9.03	8.31	7.57
TZKZ	2.85	1.51	0.86	7.23	4.74	3.76	3.02	1.59	2.30	7.86	5.19	5.20
TZKZc	1.74	1.00	1.44	9.20	8.85	8.53	3.40	0.85	1.15	8.93	8.17	7.12
DPML	3.20	2.11	1.88	7.82	5.92	5.12	3.82	1.81	2.71	8.60	6.07	5.54
DPMLc	5.71	3.76	3.14	6.99	5.76	4.83	7.43	6.48	5.59	8.15	8.02	7.32
DPMLK	2.89	1.69	1.47	7.89	5.53	4.98	3.63	1.61	2.62	8.47	5.90	5.40
DPML+	1.00	2.84	3.28	2.40	3.47	3.98	0.89	1.61	4.21	1.76	2.11	4.78
DPMV	2.13	2.45	2.42	5.90	3.84	2.94	3.64	3.09	3.41	8.04	4.94	4.34
DPMVc	3.76	3.62	4.65	4.76	4.13	4.27	5.50	4.66	5.24	7.16	5.33	5.38
DPMV+	2.22	1.71	1.69	3.07	2.22	1.86	5.22	5.31	5.60	5.95	5.76	5.86

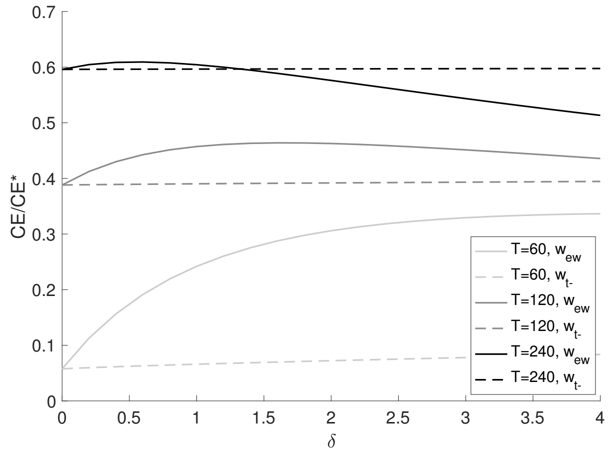
Table XII: Certainty Equivalent: Variance Targeting: Economic Cycle

This table reports the normalized CE's from variance targeting for two samples: recession months and the remaining months. The NBER recession indicator is used to identify recession months. 'Mean CE' is the mean of the normalized CE's across the datasets in Table I, and 'N(>EW)' is the number of the datasets in which a portfolio outperforms EW. δ^* indicates the models use a calibrated δ.

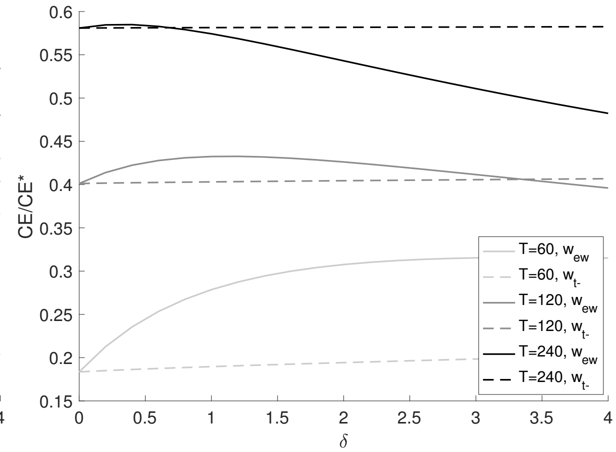
	Before Cost						After Cost					
	Mean CE			N(>EW)			Mean CE			N(>EW)		
	60	120	240	60	120	240	60	120	240	60	120	240
Recession months												
W*	1.000	1.000	1.000	12	12	12	1.000	1.000	1.000	12	12	12
EW	-0.633	-0.523	-0.457	0	0	0	-0.774	-0.646	-0.550	0	0	0
KZ	-0.462	0.329	0.086	7	12	12	-2.123	-0.304	-0.276	2	9	10
KZ(δ^*)	-0.104	0.296	0.018	11	12	12	-0.885	-0.060	-0.236	3	12	12
KZc(δ^*)	-0.458	0.271	0.052	8	12	12	-1.615	-0.218	-0.249	2	9	11
TZML	-0.198	0.256	0.037	10	12	12	-1.280	-0.227	-0.271	3	11	11
TZML(δ^*)	-0.165	0.184	-0.022	11	12	12	-0.773	-0.143	-0.266	3	11	12
TZMLc(δ^*)	-0.295	0.204	0.002	10	12	12	-1.080	-0.172	-0.253	3	11	12
TZKZ	-0.338	0.240	0.050	9	12	12	-1.702	-0.295	-0.255	2	9	11
TZKZ(δ^*)	-0.046	0.207	-0.008	12	12	12	-0.683	-0.123	-0.261	5	12	12
TZKZc(δ^*)	-0.418	0.185	0.020	8	12	12	-1.329	-0.204	-0.228	2	10	11
DPML(δ^*)	-0.088	0.117	-0.022	12	12	12	-0.673	-0.214	-0.269	5	11	11
DPMLc(δ^*)	-0.138	0.284	0.073	12	12	11	-0.692	0.062	-0.096	6	12	11
DPMLK(δ^*)	-0.088	0.079	-0.047	11	12	12	-0.666	-0.246	-0.288	6	11	12
DPML+(δ^*)	-0.363	-0.191	-0.317	10	12	12	-0.527	-0.294	-0.414	10	12	11
Remaining months												
W*	1.000	1.000	1.000	12	12	12	1.000	1.000	1.000	12	12	12
EW	0.346	0.379	0.413	0	0	0	0.347	0.392	0.433	0	0	0
KZ	0.531	0.744	0.699	9	11	9	-0.744	0.294	0.495	3	6	5
KZ(δ^*)	0.772	0.833	0.759	12	12	12	0.267	0.619	0.651	6	10	9
KZc(δ^*)	0.527	0.719	0.688	9	11	9	-0.218	0.421	0.548	5	8	7
TZML	0.706	0.773	0.696	9	10	9	-0.178	0.424	0.528	4	8	6
TZML(δ^*)	0.734	0.794	0.733	12	12	12	0.339	0.612	0.635	7	10	9
TZMLc(δ^*)	0.683	0.753	0.685	9	10	9	0.191	0.543	0.578	7	8	7
TZKZ	0.711	0.815	0.727	11	12	10	-0.320	0.454	0.572	4	8	7
TZKZ(δ^*)	0.742	0.810	0.747	12	12	12	0.363	0.634	0.642	6	10	9
TZKZc(δ^*)	0.687	0.789	0.716	10	11	10	0.161	0.579	0.619	6	9	8
DPML(δ^*)	0.718	0.798	0.729	12	12	11	0.324	0.612	0.632	7	10	9
DPMLc(δ^*)	0.451	0.537	0.564	7	8	8	0.116	0.367	0.475	7	8	8
DPMLK(δ^*)	0.725	0.801	0.725	12	12	12	0.328	0.625	0.640	7	10	9
DPML+(δ^*)	0.566	0.554	0.527	12	12	12	0.538	0.551	0.537	12	12	9

Figure 1: Optimal Deviation Penalty: No Transaction Costs

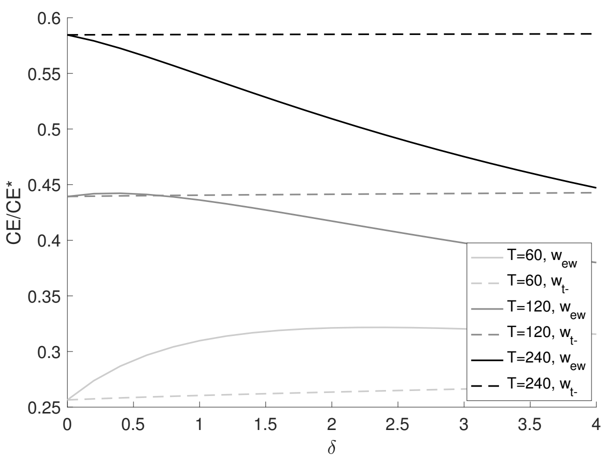
This figure displays the certainty equivalents of the four shrinkage portfolio strategies for different values of δ and estimation window sizes. Portfolios are assumed to be rebalanced monthly and managed for ten years. The vertical axis represents the normalized certainty equivalent averaged across the datasets; D1, D2, D5, and D8.



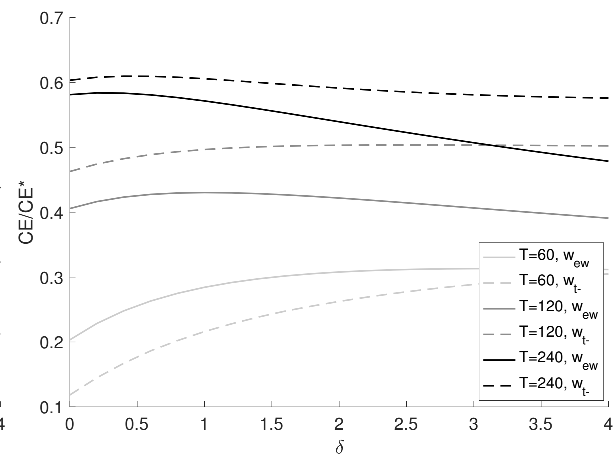
(a) KZ



(b) TZML



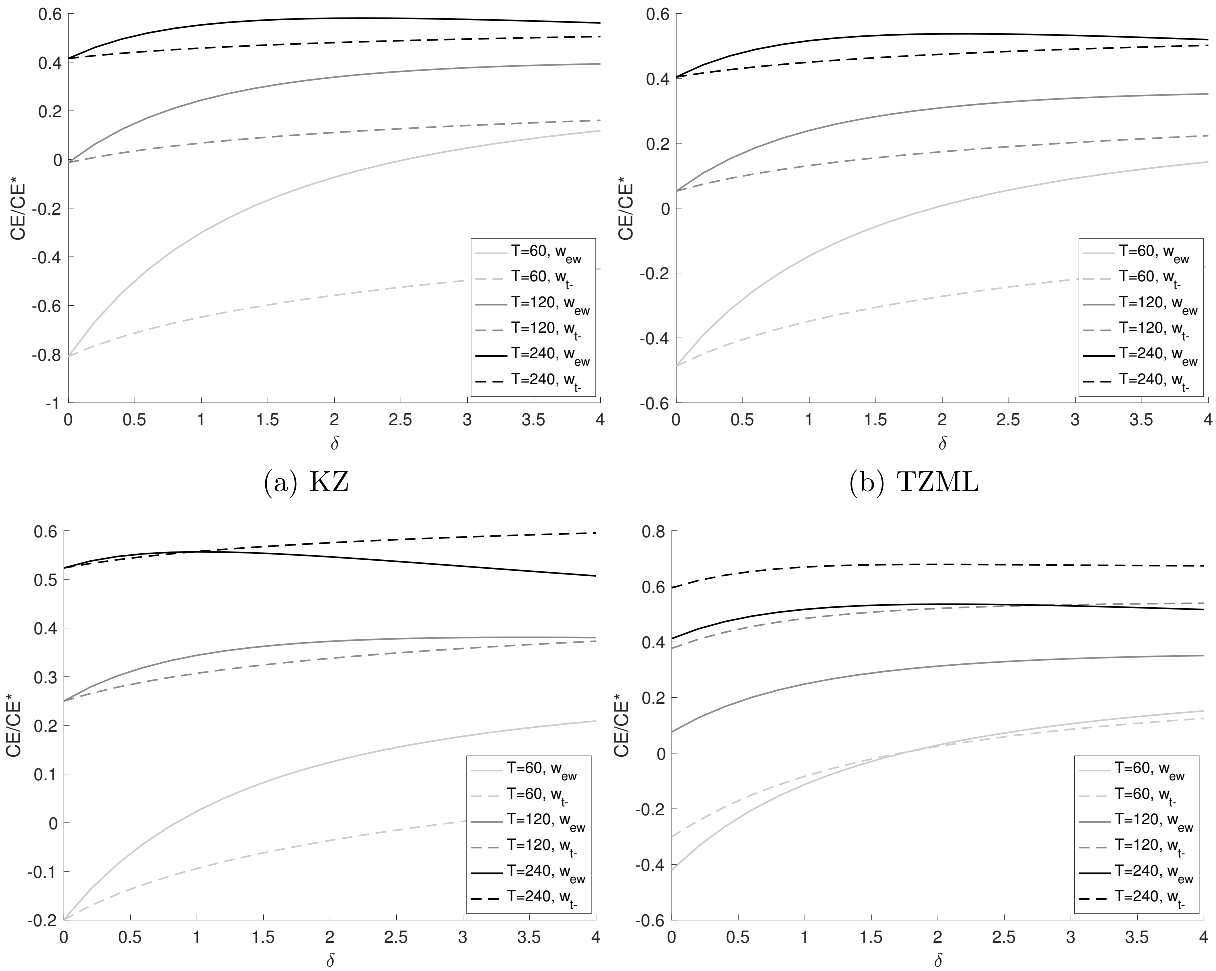
(c) TZKZ



(d) DPML

Figure 2: Optimal Deviation Penalty: 30 bp Transaction Costs

This figure displays the certainty equivalents of the four shrinkage portfolio strategies for different values of δ and estimation window sizes in the presence of 30 bp transaction costs. Portfolios are assumed to be rebalanced monthly and managed for ten years. The vertical axis represents the normalized certainty equivalent averaged across the datasets; D1, D2, D5, and D8.



(a) KZ

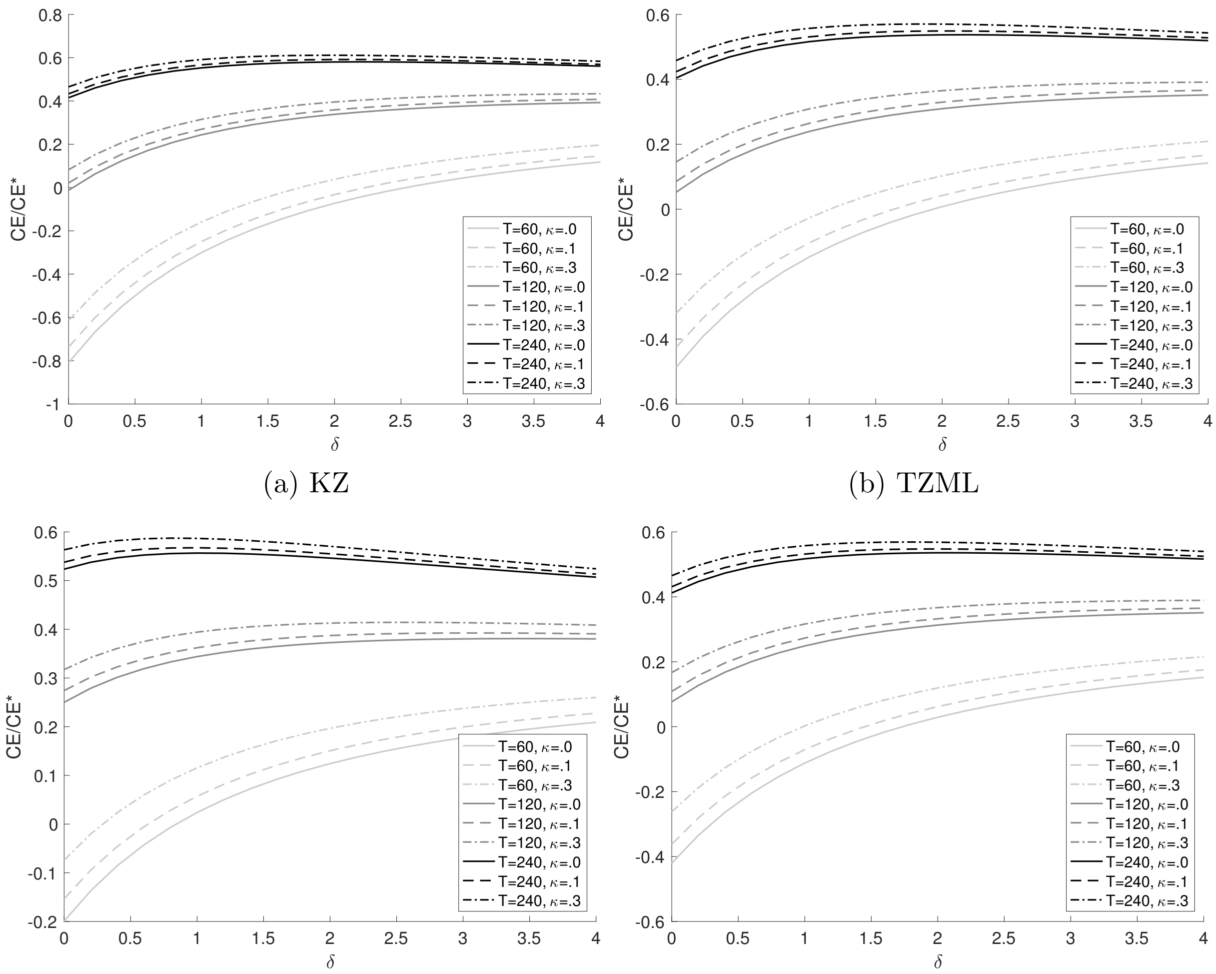
(b) TZML

(c) TZKZ

(d) DPML

Figure 3: Optimal Deviation Penalty from w_{ew} and w_{t-}: 30 bp Transaction Costs

This figure displays the certainty equivalents of the four shrinkage portfolio strategies extended using Equation (34) for different values of δ and estimation window sizes in the presence of 30 bp transaction costs. κ is the loading on w_{t-}. Portfolios are assumed to be rebalanced monthly and managed for ten years. The vertical axis represents the normalized certainty equivalent averaged across the datasets; D1, D2, D5, and D8.



(a) KZ

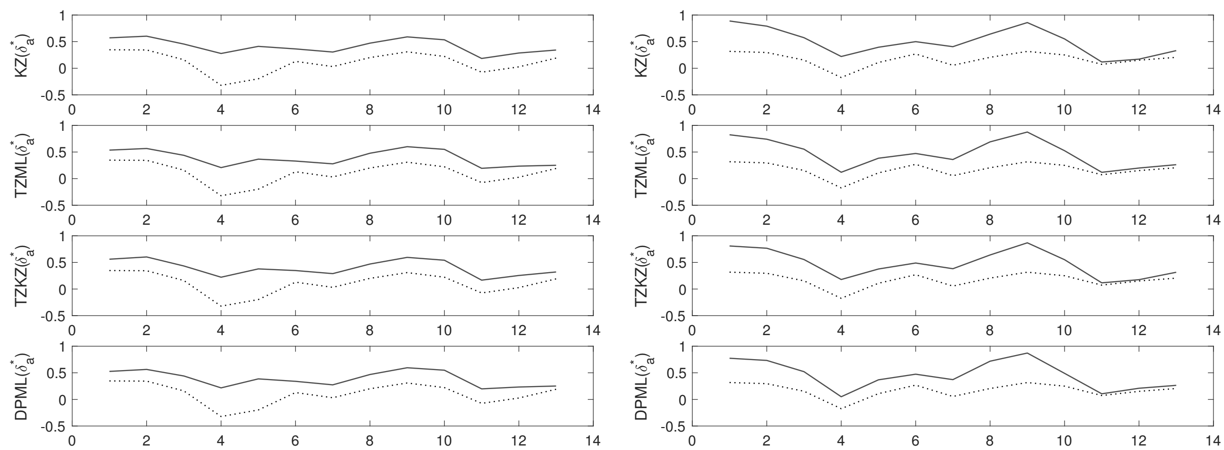
(b) TZML

(c) TZKZ

(d) DPML

Figure 4: Sub-Period Performance: CE

This figure visualizes the performance of the deviation penalty models in sub-periods (horizontal axis). Each sub-period is ten-year long and five-years apart from each other, except the last sub-period, SP13, which ends in 2015.12. Each chart displays the normalized CE's before transaction costs. The dotted lines represent EW. The results are averages across the datasets, D2-D13 (D1 is omitted due to its shorter sample period), and $T = 120$ is used.



(a) Variance Targeting (b) Utility Maximization